Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

The Economic and Statistical Value of Forecast Combinations under Regime Switching: An Application to Predictable US Returns

Massimo Guidolin
and
Carrie Fangzhou Na

Working Paper 2006-059B
http://research.stlouisfed.org/wp/2006/2006-059.pdf

November 2006
Revised April 2007

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The Economic and Statistical Value of Forecast Combinations under Regime Switching: An Application to Predictable US Returns*

Massimo Guidolin†
University of Manchester Business School and
Federal Reserve Bank of St. Louis

Carrie Fangzhou Na
Fannie Mae

Abstract

We address one interesting case – the predictability of excess US asset returns from macroeconomic factors within a flexible regime switching VAR framework – in which the presence of regimes may lead to superior forecasting performance from forecast combinations. After having documented that forecast combinations provide gains in prediction accuracy and these gains are statistically significant, we show that combinations may substantially improve portfolio selection. We find that the best-performing forecast combinations are those that either avoid estimating the pooling weights or that minimize the need for estimation. In practice, we report that the best-performing combination schemes are based on the principle of relative, past forecasting performance. The economic gains from combining forecasts in portfolio management applications appear to be large, stable over time, and robust to the introduction of realistic transaction costs.

JEL codes: C53, E44, G12, C32.

Keywords: Forecast Combination, Predictability, Multivariate Regime Switching, Portfolio Performance.

1. **Introduction**

One of the least debated stylized facts in the modern empirical finance literature is that the laws of motion governing the returns on most assets are subject to recurrent "breaks", discrete shifts in the dynamic structure generating risk premia, volatility, correlations, and – at least in principle – all relevant properties of financial returns. For instance, Paye and Timmermann (2006) find four major breaks in the past fifty years in a large set of size- and industry-sorted US equity portfolios, as well as 18 international stock market portfolios, with breaks occurring in different time periods.

Although other, competing approaches exist, one of the most widespread ways in which such evidence is translated into precise modeling strategies consists of adopting models in the regime switching class.[1] Following

*Elizabeth La Jeunesse provided excellent research assistance.We are indebted to thank Mark Wohar (a co-editor) and one anonymous referee for a number of stimulating comments and suggestions. We also thank session partecipants at the St. Louis University conference on "Forecasting in the Presence of Structural Breaks and Model Uncertainty" (August 2006).

†Correspondence to: Prof. Massimo Guidolin, Manchester Business School, Accounting & Finance Division. Phone: +44-(0)161-306-6406; Fax: +44-(0)161-275-4023. Address: University of Manchester Business School, MBS Crawford House, Booth Street East, Manchester M13 9PL, United Kingdom. E-mail: Massimo.Guidolin@mbs.ac.uk.

[1] Clements and Krolzig (1999) discuss regime switching models as ways to capture structural change.

Hamilton (1989), several papers have proposed to improve the empirical fit of standard, single-equation models for short-term interest rates (e.g., Gray (1996) and Ang and Bekaert (2002b)) and stock returns (e.g., Turner, Starz, and Nelson (1989) and Ang and Bekaert (2002a)) by allowing for mixtures of distributions. For instance, Turner, Starz, and Nelson (1989) develop a univariate model with regime shifts in means and variances, showing that mean excess equity returns tend to be low in the high-risk (volatility) period, and viceversa. Allowing for switching in the parameters of an autoregressive conditional heteroskedasticity (ARCH) process, Hamilton and Susmel (1994) report that in-sample performance and out-of-sample forecasts of the regime-switching ARCH are superior to a benchmark single-state GARCH(1,1) specification and that the high-volatility state is likely to occur in recession periods. Guidolin and Timmermann (2006a) extend this class of models to multivariate systems including excess returns on a few equity portfolios as well as bond returns.

Up to this point, the empirical finance literature has developed following an intuitive research agenda: to uncover evidence of regimes in financial data, to propose models that provide an accurate in-sample fit, and to assess their usefulness for prediction and financial decision making, e.g. portfolio choice, asset pricing, and risk management. For instance, Guidolin and Ono (2006) estimate a range of multivariate regime switching VAR models for a rich eight-variable vector that includes stock and bond returns in excess of a T-bill rate, the T-bill yield, typical predictors used in finance (such as the default spread and the dividend yield), and three macroeconomic variables: inflation, industrial production growth, and real money growth. After finding evidence of four regimes and of time-varying covariances, they show that the recursive, out-of-sample predictive performance of the four-state model is superior to a simpler (and nested) VAR(1).

However, the forecasting literature has made it clear that in the presence of structural instability in econometric relationships it is highly unlikely that a single, non-linear model – by necessity, only a misspecified simplification of the true but unknown data generating process – may provide the most accurate forecasts. On the contrary, forecasting experts have for a long time been aware that forecast combinations can be considered as a hedge against non-stationarities. For instance, Winkler (1989) argues that

> "(...) in many situations there is no such thing as a 'true' model for forecasting purposes. The world around us is continually changing, with new uncertainties replacing old ones."

Elliott and Timmermann (2004) argue that different individual forecast models may react quite differently to structural breaks, such as institutional or policy changes. Some models, such as the ones allowing for regime shifts, are only temporarily affected by the breaks and adapt more rapidly than other models. A combination that puts more weight on adaptive models before or after the breaks and loads more heavily on stable models in the middle is expected to give better results than either of the two alone. This means that the existence of structural shifts of unknown form/timing makes it likely *not* that some non-linear framework able to capture such features would manage to produce the best out-of-sample performance, but that on the contrary pooling forecasts may be the winning strategy.[2]

Our paper has two objectives. First, to present an interesting case study – the predictability of excess US asset returns from macroeconomic factors within the flexible regime switching VAR framework investigated by Guidolin and Ono (2006) – in which the presence of regimes may lead to superior forecasting performance from forecast combinations. Second – after having shown that in practice forecast combinations do provide gains

[2]For a special case, Aiolfi and Timmermann (2005) provide conditions under which the population MSFE of an equally-weighted combined forecast will be lower than the population MSFE of the best model. Clements and Hendry (1998) illustrate the usefulness of the forecast combination approach for artificial data subject to a structural break.

in prediction accuracy and that these gains are statistically significant – to investigate the economic value of such gains. We do that by assessing the improvement in (risk-adjusted) portfolio performance made available to a portfolio manager who – confronted with the evidence of regimes and the availability of a multiplicity of forecasting models – entertains the use of a number of forecast combination schemes.

Our main results can be summarized as follows. First, we extend and confirm Guidolin and Ono's (2006) findings that carefully modeling regime switches in the econometric relationships tying excess asset returns to macroeconomic and financial factors provides both a better dynamic description of the joint density of the data as well as a superior set of prediction tools.[3] Second, we show that forecast combinations may substantially improve the prediction accuracy relative to a large set of heterogeneous models. Such improvements turn out to be strong enough to carry high levels of statistical significance. Third, we find that an important qualification must be applied to our claim: the best performing forecast combinations are those that either avoid estimating the pooling weights (i.e. simple, equally weighted forecasts) or – better – those that minimize the need for estimation and, therefore, the estimation error "absorbed" by the forecaster. In practice, we report that especially at short horizons, the best performing combination schemes are based on the principle of relative, past forecasting performance (see e.g., Stock and Watson, 2003) and as such are poised to avoid estimating covariances between past forecasting performances across different models. This confirms the classical adage in forecasting that the major concern one should have about pooling forecasts involves the way to estimate the weights assigned to each individual model. A potential drawback of forecast combinations associated with this is the estimation error induced by the weight estimation. Fourth, we report large and statistically significant economic gains from combining forecasts for portfolio management purposes. In particular, we show that – depending on the risk aversion coefficient – such a value may be identified with a 8.4% increase in mean annualized portfolio returns or an increase in the realized Sharpe ratio from 0.20 to 0.36. Both measures seem hardly negligible and may represent powerful incentives for money managers to seriously entertain pooling the predictions yielded by a number of competing models as a serious alternative to simply picking one "best performing" candidate over the others. Finally, such gains turn out to be stable over time – i.e. not to depend on any speficic sub-sample of the overall 1985-2004 period – and robust to taking transaction costs into account.

Our paper contributes to two distinct literatures. On one hand, there are now numerous studies that have found that forecast combinations – even simple mean (equally weighted), trimmed mean, or median schemes – tend to outperform the best individual forecasts.[4] The general principle is that almost all econometric specifications only provide a local approximation to the real data generating process. The extreme complication in the return series makes it impossible for a single model to be superior to others all the time. If the information incorporated in two (sets of) models are not completely overlapping, the combination of the two may give better forecasts than either alone. These ideas are particularly relevant in the presence of non-stationarities, which are notoriously difficult to effectively model without recurring to approximations.

On the other hand, an impressive literature has piled up that investigates whether US stock and bond returns are predictable using past values of macroeconomic variables. In fact, using regression models, numer-

[3]Notice that in a forecasting perspective there appears to be no clear consensus as to whether allowing for non-linearities may concretely lead to an improved forecasting performance (see e.g. De Gooijer and Kumar, 1992, Clements and Krolzig, 1998). Stock and Watson (2001) report that non-linear models tend to outperform linear ones only at short horizons. The gains are however small.

[4]General discussion of forecast combinations and additional references may be found in Clemen (1989), Stock and Watson (2001), Diebold and Lopez (1996), Hendry and Clements (2002), and Timmermann (2005).

ous studies have found that in each data set a few macroeconomic variables can be found that systematically predict US stock returns. Among others, Fama and French (1988) document that the dividend yield forecasts future returns on common stocks. Fama and Schwert (1977) report that real stock returns are negatively related with expected and unexpected components of inflation, and that industrial production and real GNP growth have forecasting power for financial returns. Campbell (1987) presents evidence that a variety of term structure variables such as two- and six-month spreads as well as the 1-month T-bill rate, all forecast excess stock and bond returns. Fama and French (1989) confirm this result using data at alternative frequencies and a longer sample period (1927-1987). Similarly, Fama and French (1989) show that the default spread is a significant predictor of stock returns. Fama and French (1993) extend this evidence to excess bond returns.

A related paper is Guidolin and Timmermann (2005b), in which optimal combination weights are derived within a four-state regime switching model that captures common latent factors driving short-term US spot and forward rates. In their framework, the combination weights explicitly depend on the state of the economy. They find that combining forecasts from different models helps improve the out-of-sample forecasting performance for short-term interest rates in the US. Apart from obvious differences in the application, our paper differs from theirs because it does not impose the presence of regimes when computing weights, but simply uses realized forecast errors to uncover the optimizing weights. Moreover, while Guidolin and Timmermann (2005b) refrain from an assessment of the economic value of their regime-switching forecast combinations, in our paper we attach a precise price tag to the usefulness of combinations in portfolio management.

The paper is structured as follows. Section 2 describes the forecasting models. Section 3 gives information on the data employed in the paper. Section 4 estimates a range of regime switching models and proceeds to select the one providing the best fit according to a number of statistical criteria. Parameter estimates and interpretation for the resulting regimes are provided. Section 5 shows that a four-state model produces useful out-of-sample forecasts. Section 6 introduces the forecast combination problem, comments on the recursive combination weights derived applying standard methods, and gives results on the relative performance of combinations in out-of-sample experiments. Section 7 evaluates the economic value of forecast combinations by focussing on a real time, pseudo out-of-sample portfolio choice experiment. Section 8 concludes.

2. The Models

Suppose that the random vector collecting monthly (excess) returns on l different assets (the vector $\mathbf{x}_t$) and q macroeconomic variables ($\mathbf{m}_t$), possibly predicting (and predicted by) asset returns, follows a $k-$regime Markov switching (MS) $VAR(p)$ process with heteroskedastic components, compactly $MSIAH(k,p)$ (see Krolzig, 1997):[5]

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{j,S_t}\mathbf{y}_{t-j} + \boldsymbol{\Sigma}_{S_t}\boldsymbol{\epsilon}_t \tag{1}$$

with $\boldsymbol{\epsilon}_t \sim NID(\mathbf{0}, \mathbf{I}_{l+q})$ and $\mathbf{y}_t \equiv (\mathbf{x}_t' \ \mathbf{m}_t')'$.[6] S_t is a latent state variable driving all the parameters in equation (1). $\boldsymbol{\mu}_{S_t}$ collects the $l+q$ regime-dependent intercepts, while the $(l+q)\times(l+q)$ matrix $\boldsymbol{\Sigma}_{S_t}$ represents the factor in a state-dependent Choleski factorization of the covariance matrix, $\boldsymbol{\Omega}_{S_t}$. A non-diagonal $\boldsymbol{\Sigma}_{S_t}$ captures *simultaneous* co-movements. Moreover, *dynamic* (lagged) linkages across different asset markets and

[5] I, A and H refer to state dependence in the intercept, vector autoregressive terms and covariance matrix. p is the autoregressive order. Models in the class MSIH$(k,0)$-$VAR(p)$ have regime switching in the intercept but not in the VAR coefficients.

[6] Assume the absence of roots outside the unit circle, thus making the process covariance stationary. Ang and Bekaert (2002b) show that for covariance stationarity to obtain, it is sufficient for such a condition to be verified in at least one of the k regimes.

between financial markets and macroeconomic influences are captured by the VAR(p). In fact, conditionally on the unobservable state S_t, (1) defines a standard Gaussian, reduced-form VAR(p) model. On the other hand, when $k > 1$, alternative hidden states are possible that will influence both the conditional mean and the volatility/correlation structure characterizing (1). The states are generated by a discrete, homogeneous, irreducible and ergodic first-order Markov chain:[7]

$$\Pr\left(S_t = j | \{S_j\}_{j=1}^{t-1}, \{\mathbf{y}_j\}_{j=1}^{t-1}\right) = \Pr\left(S_t = j | S_{t-1} = i\right) = p_{ij}, \tag{2}$$

where p_{ij} is the generic $[i, j]$ element of the $k \times k$ transition matrix $\mathbf{P}$. Ergodicity implies the existence of a stationary vector of probabilities $\bar{\boldsymbol{\xi}}$ satisfying $\bar{\boldsymbol{\xi}} = \mathbf{P}'\bar{\boldsymbol{\xi}}$.

(1) nests a number of simpler model in which either some parameter matrices are not needed or become independent of the regime. The simpler models may greatly reduce the number of parameters to be estimated. Among them, we will devote special attention to $MSIH(k,0)$-$VAR(p)$ models,

$$\mathbf{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \boldsymbol{\Sigma}_{S_t} \boldsymbol{\epsilon}_t, \tag{3}$$

a special case of equation (1) in which intercepts and covariance matrix are regime-dependent, while the VAR(p) coefficients are not. This restricted sub-class of models turns out to be important in the following. A limit case of (1) is obtained when $k = 1$, a standard multivariate Gaussian VAR(p) model:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{t-j} + \boldsymbol{\Sigma} \boldsymbol{\epsilon}_t. \tag{4}$$

Three additional models are entertained in the following: Pesaran and Timmermann (1995)-style predictive regressions, Box-Jenkins style univariate ARIMA models, and simple "no-change" benchmarks (see e.g., Neely and Weller, 2000) for excess asset returns. Pesaran and Timmermann's (PT) regressions imply that at time t

$$\mathbf{x}_t = \boldsymbol{\mu}^x + \sum_{j=1}^{p^x} \mathbf{B}_j \mathbf{y}_{t-j} + \boldsymbol{\Sigma}^x \mathbf{u}_t \quad \mathbf{u}_t \sim NID(\mathbf{0}, \mathbf{I}_l), \tag{5}$$

i.e. past values of both excess returns and of macroeconomic variables predict subsequent excess returns. Taken at face value, equation (5) simply corresponds to the first l rows of equation (4). In fact, Pesaran and Timmermann (1995) experiment with a number of statistical criteria to recursively select which of the variables in $\mathbf{y}_t$ ought to be included in the predictive regression. In what follows we experiment with two such criteria, the selection of variables from $\mathbf{y}_t$ which either (i) maximize the adjusted $\bar{R}^2$, or (ii) minimize the BIC criterion.

As for the univariate ARIMA class, we restrict ourselves to consider Gaussian $AR(p)$ models, which are special cases of (5),

$$x_t^n = \phi_0^n + \sum_{j=1}^{p^n} \phi_j^n x_{t-j}^n + \sigma^n u_t^n \quad u_t^n \sim NID(0, 1), \tag{6}$$

[7]The assumption of a first-order Markov process is not restrictive, since a higher order Markov chain can always be reparameterized as a higher dimensional first-order Markov chain. On the other hand, relaxing the assumption of fixed transition probabilities (see, e.g., Diebold, Rudebusch, and Sichel, 1993) complicates estimation but could lead to improved forecasting performance. We leave this extension for future research.

$(n = 1, ..., l)$ in which restrictions are imposed on the coefficient matrices $\mathbf{B}_j$. The number of lags is selected by minimizing the BIC criterion.[8]

The no change benchmarks correspond to (5) when $\mathbf{B}_j = \mathbf{O}$ for $j = 1, ..., p^x$, i.e. when $\mathbf{x}_t = \boldsymbol{\mu}^x + \boldsymbol{\Sigma}^x \mathbf{u}_t$ which implies the following (approximate) Gaussian random walk with drift for log-asset prices:

$$ln(\mathbf{P}_t) = ln(\mathbf{P}_{t-1}) + r^f \boldsymbol{\iota}_l + \boldsymbol{\mu}^x + \boldsymbol{\Sigma}^x \mathbf{u}_t = \mathbf{d} + ln(\mathbf{P}_{t-1}) + \boldsymbol{\Sigma}^x \mathbf{u}_t \quad (7)$$

which is derived from the definition $\mathbf{x}_t \equiv ln(\mathbf{P}_t) - ln(\mathbf{P}_{t-1}) - r^f \boldsymbol{\iota}_l$ where r^f is the risk-free interest rate and $\boldsymbol{\iota}_l$ is a $l \times 1$ column vector of ones. In this case, $\boldsymbol{\mu}^x$ is the risk premium, assumed to be constant over time. Notice that in principle, the criteria of variable selection within PT's framework may endogenously yield (7), although in general we would expect this not to be the case. equation (7) has played a key role in the development of the modern theory of asset pricing and portfolio choice, while it also corresponds to a natural random walk yardstick popular in forecasting and applied time series analysis. Therefore it seems natural to compare our models – and forecast combinations built on them – to this constant risk premium benchmark.[9]

3. The Data

We use monthly data for the period 1926:12 - 2004:12, a total of 937 observations per variable. Asset returns are from the Center for Research on Security Prices (CRSP) at the University of Chicago. In particular, we use data on the three most important segments of the US market: stocks (value-weighted stock returns for the NYSE, NASDAQ, and the AMEX exchanges), bonds (a CRSP index of 10-year to maturity US government bonds), and money market instruments (30-day Treasury bills, again from CRSP).[10]

Additionally, we employ five predictor variables which either correspond to macroeconomic aggregates or that have been associated to business cycle conditions in earlier research. In the first group we have the CPI inflation rate (seasonally adjusted), the rate of growth of industrial production (seasonally adjusted), and the rate of growth of a measure of adjusted monetary base. These three series are available from FRED® II at the Federal Reserve Bank of St. Louis. The practice of seasonally adjusting the data in real time experiments (See sections 5-7) assumes that market participants are effectively able to 'see through' the veil of time series variation purely caused by seasonal factors. In the latter group we have two variables. The first is the dividend yield (calculated from CRSP data), defined as aggregate dividends on the value-weighted CRSP portfolio of stocks over the previous twelve month period divided by the current stock price. The second is the default spread, defined as the differential yield on Moody's Bbb (low rating) and Aaa (high rating) seasoned corporate bonds with similar maturities. These two variables have played a key role in the recent literature on optimal asset allocation under predictable asset returns, see e.g. Fugazza, Guidolin, and Nicodano (2006).

In our empirical analysis we study *excess* stock returns, defined as the difference between nominal, realized monthly returns and the 1-month T-bill rate. For similar reasons, given the important literature on term spreads in the US yield curve, we use the long-short bond term spread (the term premium) defined as the excess return of CRSP long-term bond over 1-month T-bill returns. Finally, also the monetary base growth is measured in real terms, by deducting from nominal rates of growth the realized CPI inflation rate.

[8]Notice that settling on some value $p^n \geq 2$ does not impose that $\phi_j^n \neq 0$ for $j < p^n$. For instance a model such as $x_t^n = \phi_0^n + \phi_1^n x_{t-1}^n + \phi_3^n x_{t-3}^n + \sigma^n u_t^n$ might be selected.

[9]Coulson and Robins (1993) advocate using "no-change" (random walk) models in forecast combination experiments.

[10]The bond returns data are completed by using the Ibbotson-Sinquifeld data for the period 1926-1946.

Tables 1 reports basic summary statistics. Mean values are consistent with commonly known facts: for instance, the mean excess stock return is 0.65% per month, i.e. 7.80% per year, which represents a typical value in the equity premium literature, with an annualized volatility of 19.1%; the mean term premium is 0.14% per month, i.e. 1.68% per year, a moderate but plausible average slope of the US term structure; the average annualized *real* T-bill rate is 0.60% which, summed to a mean annualized inflation rate of 3.12%, delivers a mean annualized nominal short-term interest rate of 3.72%, once more in line with the typical values reported in the asset pricing literature. Both real money and industrial production growth are positive on average, 0.48 and 2.52 percent in annualized terms, respectively.

Table 1 approximately here

All the series display evident departures from a (marginal) Gaussian distribution, which would imply zero skewness (i.e. a symmetric distribution) and a kurtosis coefficient of 3. On the opposite, both excess stock returns and all macroeconomic variables are characterized by huge kurtosis values (in excess of 10), an indication of distributions with tails considerably fatter than a normal. The dividend yield has only moderate kurtosis, but it is also skewed to the right (which is to be expected, since the dividend yield cannot be negative by construction). Even in the case of excess bond returns, a formal Jarque-Bera test of marginal normal distribution rejects with a 0.000 p-value. For all series (the exception is excess bond returns, which are not serially correlated) there is evidence of statistically significant first-order serial correlation, as evidenced by Ljung-Box statistics (of order 4) in excess of the 1% critical value under a $\chi^2_{(4)}$. Similarly, there is evidence of volatility clustering (heteroskedasticity), as all Ljung-Box statistics (of order 4) applied to squared values of the variables are highly significant.

4. **Econometric Estimates**

This section reviews some of the results in Guidolin and Ono (2006) and contains the bulk of our estimation, in-sample results. Section 4.1 presents a number of model selection criteria and shows that a four-state model in which the VAR component of the model is time-homogeneous outperforms a number of competing models in many dimensions. Section 4.2 presents parameter estimates for such a model. The sub-section establishes that a regime switching model is required to provide an accurate in-sample fit to the data.

4.1. *Model Selection*

We estimate a large number of variants of (1) and use five alternative criteria to gauge of the correct specification of the candidate models. In particular, we estimate models with $k = 1, 2, 3, 4$, with and without regime-dependent covariance matrix, and with VAR coefficients that may or may not depend on the regime.[11]

The first selection criterion concerns the appropriate number of regimes k in model (1). In particular, we would like to test whether the null of a single-state model ($k = 1$) can be rejected in favor of $k > 1$. As already stressed, when $k = 1$ equation (1) reduces to a simpler Gaussian VAR(p) model. As discussed in Garcia (1998), testing for the number of states in a regime switching framework may be tricky. Given some $k \geq 2$, the problem is that under any number of regimes smaller than k there are a few parameters of

[11] We refrain from trying and estimating models with $k \geq 5$ since the number of parameters quickly grows to levels that make either estimation uncertainty overwhelming or that cause the MLE-EM routines to fail. For instance, a MSIAH(5,1) model implies 560 paramters, i.e. only 13.4 observations per parameter.

the unrestricted model – e.g., some (or all) elements of the transition probability matrix associated to the rows that correspond to "disappearing states" — that can take any values without influencing the likelihood function. We say that these parameters become a nuisance to the estimation. The result is that the presence of nuisance parameters gives the likelihood surface so many degrees of freedom that computationally one can never reject the null that the non-zero values of those parameters were purely due to sampling variation. This implies that asymptotically the LR statistic fails to have a standard chi-square distribution. Davies (1977) circumvents the nuisance parameters problem by deriving an upper bound for the significance level of the LR test under nuisance parameters:

$$\Pr\left(LR > x\right) \leq \Pr\left(\chi_1^2 > x\right) + \sqrt{2x}\exp\left(-\frac{x}{2}\right)\left[\Gamma\left(\frac{1}{2}\right)\right]^{-1}.$$

where $\Gamma(\cdot)$ is the standard gamma function. We find that even adjusting for the presence of nuisance parameters, the evidence against specifying traditional single-state models is overwhelming: the smallest LR statistic takes a value of 139, which is clearly above any conceivable critical value regardless of number of restrictions imposed.

Once we establish that $k \geq 2$ is appropriate, this only rules out models of simpler VAR type. We therefore proceed to select an appropriate model within the more general regime switching class $MSIAH(k,p)$ with $k \geq 2$. As in a few other applied papers on regime switching models (e.g., Sola and Driffill (1994) and Guidolin and Timmermann (2006c)), we employ a battery of information criteria, i.e. the Akaike (AIC), Bayes-Schwartz (BIC), and Hannan-Quinn (H-Q) criteria. These criteria are supposed to trade-off in-sample fit with prediction accuracy. In practice, information criteria identify the ex-ante potential for good out-of-sample performance by penalizing models with a large number of parameters. We find evidence of some tensions among different criteria. The AIC is minimized by a richly parameterized MSIAH(4,1) model in which 444 parameters have to be estimated. Although MLE estimation could be carried out, issues may exist with a model that implies a saturation ratio (i.e. the number of available observations per estimated parameter) of only 16.9. However, this is less than surprising as the AIC is generally known to select large models in nonlinear frameworks (see e.g., Fenton and Gallant (1996)). Next, the H-Q seems to be undecided between a relatively parsimonious MSIH(4,0)-VAR(1) model (with saturation ratio of almost 30) and a richer MSIAH(3,1) (saturation ratio of 23). Notice that these two models imply a different number of regimes, 3 vs. 4. So, if on one hand it seems obvious that regime switching matters, the precise number of states required remains debatable. Finally, the BIC selects a relatively tight MSIH(4,0)-VAR(1) model.

All in all, we are left with two plausible and competing candidate models. The first one is a four-regime MSIH(4,0)-VAR(1) model that is directly selected by both the H-Q and the parsimonious BIC criterion. The second is a three-regime MSIAH(3,1) model that obtains a good 'score' in a H-Q metric.[12] As discussed by Guidolin and Ono (2006), these two models are structurally different both in a statistical and in an economic sense. In particular, MSIAH(3,1) implies regime switching in the VAR(1) coefficients, i.e. in this model the dynamic linkages between financial markets and the macroeconomy are time-varying; this is not the case for the MSIH(4,0)-VAR(1) model.

Finally, density specification tests are used to "break the impasse" between MSIAH(3,1) and MSIH(4,0)-VAR(1). Regime switching models consists of flexible mixtures which – if the number of regimes k is expanded

[12] We do not pursue estimation of the richer MSIAH(4,1) (selected by the AIC) because of the high probability of it being over-parameterized (its saturation ratio is half the MSIH(4,0)-VAR(1)).

with the sample size $-$ may be thought of as providing a seminonparametric approximation of the process followed by the joint conditional density of the data, see Marron and Wand (1992). In this framework, it has become customary to require of a regime switching model that they provide a correct specification for the entire conditional distribution of the variables at hand. The seminal work of Diebold et al. (1998) has spurred increasing interest in specification tests based on the h-step ahead accuracy of fit of a model for the underlying density. These tests are based on the probability integral transform or z-score. This is the probability of observing a value smaller than or equal to the realization of $\mathbf{y}_{t+1}$, $\tilde{\mathbf{y}}_{t+1}$, under the null that the model is correctly specified. Under a k-regime mixture of normals, this is given by $(n = 1, ..., l + m)$

$$\Pr\left(y_{t+1}^n \leq \tilde{y}_{t+1}^n | \Im_t\right) = \sum_{i=1}^{k} \Phi\left(\sigma_{n,i}^{-1}\left[y_{t+1}^n - \mu_{n,i} - \mathbf{e}_n' \sum_{j=1}^{p} \mathbf{A}_{j,i}\mathbf{y}_{t+1-j}\right]\right) \Pr(S_{t+1} = i|\Im_t) \equiv z_{t+1}^n,$$

where $\sigma_{n,i}$ is the volatility of variable n in state i, and $\mathbf{e}_n$ is a vector with a one in position n and zeros elsewhere. If the model is correctly specified, z_{t+1}^n should be independently and identically distributed (IID) and uniform on the interval $[0, 1]$.[13]

Unfortunately, testing whether a distribution is uniform is not a simple task, as tests popular in the statistics literature often rely on the IID-ness of the series, which is here at stake as well. Therefore Berkowitz (2001) has recently proposed a likelihood-ratio test that inverts Φ to get a transformed z-score (dropping the n superscript),

$$z_{t+1}^* \equiv \Phi^{-1}(z_{t+1}),$$

which essentially transforms the z-score back into a bell-shaped random variable. Provided that the model is correctly specified, z^* should be IID and normally distributed (NID(0, 1)). To test this hypothesis, we use a likelihood ratio test that focuses on a few salient moments of the return distribution. Suppose the log-likelihood function is evaluated under the null that $z_{t+1}^*\tilde{}\ NID(0, 1)$:

$$L_{IIN(0,1)} \equiv -\frac{T}{2}\ln(2\pi) - \sum_{t=1}^{T} \frac{(z_t^*)^2}{2}.$$

Under the alternative of a misspecified model, the log-likelihood function incorporates deviations from the null, $z_{t+1}^*\tilde{}\ NID(0, 1)$:

$$z_{t+1}^* = \alpha + \sum_{j=1}^{q}\sum_{i=1}^{l} \beta_{ji}(z_{t+1-i}^*)^j + \sigma u_{t+1}, \tag{8}$$

where $u_{t+1} \sim NID(0, 1)$. The null of a correct return model implies $q \times l + 2$ restrictions – i.e., $\alpha = \beta_{ji} = 0$ ($j = 1, ..., q$ and $i = 1, ..., l$) and $\sigma = 1$ – in equation (8). Let $L(\hat{\alpha},\ \{\hat{\beta}_{ji}\}_{j=1}^{q}\,{}_{i=1}^{l},\ \hat{\sigma})$ be the maximized log-likelihood obtained from equation (8). To test that the null model (some version of (1)) is correctly specified, we can then use the following test statistic:

$$LR_{ql+2} \equiv -2\left[L_{IIN(0,1)} - L(\hat{\alpha}, \{\hat{\beta}_{ji}\}_{j=1}^{q}\,{}_{i=1}^{l}, \hat{\sigma})\right] \xrightarrow{d} \chi_{ql+2}^2.$$

In addition to the standard Jarque-Bera test that considers skew and kurtosis in the z-scores to detect non normalities in z_{t+1}^*, it is customary to present three likelihood ratio tests, namely a test of zero-mean and

[13] The uniform requirement relates to the fact that deviations between realized values and projected (fitted) ones should be conditionally normal and as such describe a uniform distribution once it 'filtered through' an appropriate Gaussian cdf. The IID condition reflects the fact that if the model is correctly specified, errors ought to be unpredictable and fail to show any detectable structure.

unit variance ($q = l = 0$), a test of lack of serial correlation in the z-scores ($q = 1$ and $l = 1$) and a test that further restricts their squared values to be serially uncorrelated in order to test for omitted volatility dynamics ($q = 2$ and $l = 2$). Notice that a rejection of the null of normal transformed z-scores has the same meaning as rejecting the null of a uniform distribution for the raw z-scores, i.e. the model fails in generating a density with the appropriate shape. A rejection of the zero-mean, unit variance restriction points to specific problems in the location and scale of the density underlying the model. A rejection of the restriction that $\{z^*_{t+1}\}$ is IID (i.e. the presence of serial correlation in levels or squares) points to dynamic misspecifications.

Strikingly, a simple and yet popular VAR(1) is resoundingly rejected by *all* tests and for *all* variables. Rejections tend to be harsh: the highest VAR(1) p-value is 0.001, i.e. there is actually a very thin chance that the data might have been generated by a simple linear Gaussian homoskedastic model. In fact, the rejections are so strong that it becomes difficult to understand in which direction one should be moving to amend the VAR(1) model to improve the in-sample performance.[14]

The picture improves, albeit not drastically, when a MSIAH(3,1) model is estimated. For most tests and variables, the LR test statistics decline by a factor between 30 and 200% when we move from a single- to a multi-state model. The exceptions are few (the Jarque-Bera tests for excess bond returns, the inflation rate, and adjusted monetary base growth). However. all (but one, the real T-bill rate and when testing the zero-mean unit-variance properties) of the related p-values remain highly significant, indicating strong rejection of the null of correct specification of the three-state model.

Figure 1 provides visual evidence on the sources of misspecifications within a MSIAH(3,1) model by displaying quantile-quantile (q-q) plots for the empirical distribution of $\{z^*_{t+1}\}$ for each of the eight variables. Notice that if $\{z^*_{t+1}\}$ is $N(0,1)$, the q-q should approximately look like a 45 degrees straight line in the q-q plane. This seems to happen only for real T-bill yields. On the other hand, a few plots assume an S-shape, i.e. the slope is too high at the center of the distribution (i.e. more mass is put under the distribution of $\{z^*_{t+1}\}$ than under a $N(0,1)$) and too flat for intermediate values in the support (where mass is missing vs. the Gaussian case). This is the case of excess stock and bond returns, inflation, IP and monetary growth rates. In two other cases − dividend yield and the default spread − the q-q plots are flatter than a 45 degrees line, a sign that too much mass is simply moved to the tails of the corresponding distributions.

Figure 1 approximately here

On the contrary, the improvement is strong and significant when we fit a four-state model. The p-values associated to the various tests generally increase and out of 32 combinations test/variable, we have that the null of no misspecification fails to be rejected in 15 cases, with p-values exceeding 0.05. Of the remaining 17 tests, in 7 the p-values are between 0.01 and 0.05, i.e. the rejection is rather mild. For 3 variables − remarkably all of the financial variables, including the dividend yield − the tests give evidence of correct specification, with only some concerns caused by the potential of additional volatility clustering not accommodated by regime switching covariance matrices in excess stock returns, and by deviations of the shape of the bond and dividend yield scores from normality. The improvement vs. the three-state model is clear: in only one case the LR statistic increases when the number of regimes is increased (the test for correct location and scale of the distribution of the transformed z-scores) and the variation in the corresponding p-value is moderate. Figure 2 presents almost perfect q-q plots, i.e. roughly aligned around a 45 degrees line.

[14] Detailed results are reported in Guidolin and Ono (2006).

Figure 2 approximately here

4.2. *A Four-State Model*

Table 2 presents parameter estimates. Panel A reports estimates of a benchmark, single-state VAR(1) model. Panel B shows MLE-EM estimates of the four-state model. Panel A shows that in a standard VAR(1) many of the estimated coefficients are not significant. The implications for predictability of financial returns are rather interesting: apart from a weak own serial correlation (coefficient is 0.10), excess stock returns are essentially unpredictable using any of the macro instruments entertained in the paper. The same is true for excess bond returns, which can be just (weakly) predicted from past excess stock returns (coefficient -0.02). Much more predictability characterizes real short-term interest rates, which are (as expected) highly persistent (coefficient 0.58) and can also be predicted off past default spreads (coefficient 1.28) and IP growth (-0.06). However, only limited confidence should be attributed to these results for three reasons: we know from Section 4.1 that single-state VAR(1) models are rejected even when account is taken of nuisance parameter issues. If there are multiple regimes in the data, we can expect that all estimates obtained from single-regime models might be severely biased.[15]

Table 2 approximately here

Things greatly improve under a well-specified regime switching model. Table 2, panel B, starts by showing that the fraction of parameters in the conditional mean function that are precisely estimated grows when multiple states are allowed. For instance, most of the intercepts parameters are now highly significant. However the most visible changes concern indeed the amount (and in some cases, the structure) of the predictability patterns implied by the model. On one hand, modeling regimes erases all traces of own- and cross-serial correlation involving excess asset returns. This is not surprising as structural breaks (regimes) are well known to artificially inflate the degree of own-persistence. On the other hand, excess stock and bond returns become now highly predictable using lagged values of three variables: real T-bill rates (which forecast lower excess returns, since the real short-term rate enters the discount rate in asset pricing models) as in Campbell (1987), the default spread (which forecasts higher excess returns, as in Fama and French (1989)), and the inflation rate (which forecasts lower future excess returns, presumably as a consequence of the recessions that need to be induced to bring inflation under control) as in Fama and Schwert (1977). Excess stock returns are also predicted by past IP growth (as in Cutler, Poterba and Summers (1989)), although the economic effect is small.[16]

Table 2 also reports the regime-dependent estimated volatilities and pairwise correlations implied by estimated variances and covariances. With limited exceptions, regimes 1 and 2 are characterized by moderate

[15] Guidolin and Ono (2006) also show that the fit provided by a VAR(1) model is rather poor. For instance, the VAR does match the data sample means, i.e. over the long run the VAR(1) forecasts values for the variables that are often radically different from those observed on average in-sample.

[16] We also obtain evidence of predictability of macro variables, especially IP growth which is not only persistent, but also predicted by past T-bill yields and default spreads with significant coefficients, similarly to Stock and Watson (1989). Interestingly, the ability of asset returns to predict macroeconomic conditions − principally future real growth − seems confined to single-regime models, see e.g. James, Koreisha, and Partch (1985). When $k = 4$ most coefficients lose significance. This means that the bulk of predictability for inflation and real growth comes from the regime switching structure (and past macroeconomic conditions), and not from financial markets.

volatilities of the shocks and by correlations which tend to be smaller (in absolute value) than in the single-state VAR(1) model of Panel A. On the opposite, regimes 3 and 4 imply higher volatilities and (at least for a majority of pairs) larger correlations in absolute value. In fact, Table 2 and Figure 3 help us giving some economic interpretation to the four regimes. Regime 1 is a bull/rebound state characterized by high equity risk premia (14.5% on annualized basis), low or negative real short term interest rates, relatively high inflation (4.6% on annual basis), and high dividend yields. In this regime, all variables display moderate volatility, e.g. 13% for excess equity returns, 2.4% for excess bond returns, and 2.5% for inflation. This is a rebound state because its persistence is moderate (approximately 10 months) and it tends to follow bear regimes: the estimated transition matrix in Table 2 shows that starting from a bear/recession regime, 17% of the time the system accesses a rebound (76% it stays in a bear regime). As a result, the mean dividend yield tends to be exceptionally high (5.2% vs. a historical mean of 3.8%), an indication of the existence of good bargains in the stock market. The exceptional stock market performance tends to be disjoint from real growth, which has actually an unconditional mean of only 0.84% per annum. Consistently, the yield curve is flat (the annualized term premium is 1.9%). Historically this regime coincides with the stock market bubble of 1927-1929, the Great-Depression rebound of 1934-1937, and most of the WWII and immediate post-war years. After one spike in the mid-1950s, the occurrences of this regime have been rather episodic, although some late periods in the tech bubble of 1999-2000 are captured by this state.

Figure 3 approximately here

Regime 2 is a stable (low volatility) regime characterized by good real growth (2.9% per year) and moderate inflation (3.2%). This a persistent regime (15 months on average) in which also equity risk premia are fairly high (5.4%), although equity prices correspond to much higher multiples than in regime one (the dividend yield has unconditional mean of 2.8%, below the historical sample mean). As experienced in the 1990s, real short term rates are low and default-free credit cheap, just in excess of 1% per year. In fact, regime 2 captures most of the booming years between the mid 1950s and 1974 (the interruptions correspond to officially dated NBER recessions, picked up by state 4). The 1990s as well as the more recent, 2002-2004 period are entirely captured by regime 2.

Regime 3 describes periods of intense real growth, the initial stages of the business cycle when the economy emerges from a trough. In fact, regime 3 is scarcely persistent (5 months on average). This state is characterized by high real growth and equity premia (5.9%), and an upward sloping yield curve. Figure 3 shows that − among other periods − the early 1980s and 1990s are captured by regime 3. Finally, regime 4 represents a bear/recession state, in which risk premia are small, dividend yields relatively high (as stock prices decline), and inflation and real growth are both negative. Consistently with this interpretation, the default spread is high in this state (25 basis points vs. a historical mean of 9 points only), while its duration is moderate (4 months), which reflects the fact that recession and bear markets are generally short-lived. This state also implies high volatility. Figure 3 shows that regime 4 picks up all major US recessions after WWII, in addition to a long period that matches the so-called Great Depression.

Figure 4 reinforces the impression that the four-state model provides an accurate fit to the data by plotting the in-sample fitted values for the eight series under examination. On the top of each panel, correlations between fitted and observed values are reported. First, we notice that in many instances, the MSIH(4,0)-VAR(1) values are simply much more volatile (hence able to track the underlying series) than VAR(1) fit, which is to be expected. A careful eye may even detect the existence of periods (e.g., the 1990s) over which the behavior of the fit values across the two models is clearly heterogeneous. Second, the four-state model does a superior job at matching the dynamics of most of the series: for six out of eight, the correlation between actual values and fitted ones is higher under a MSIH(4,0)-VAR(1) than under a VAR(1).

Figure 4 approximately here

5. **Forecasting Performance**

As a first step towards investigating the value of forecast combinations under regime switching, in this Section we proceed to establish that a four-state MSIH(4,0)-VAR(1) also precisely predicts most of the variables of interest, especially excess asset returns. In this Section, a Reader will also notice that our focus gradually shifts from the overall vector $\mathbf{y}_t$ to $\mathbf{x}_t$, the set of excess stock and bond returns. The role of other variables – especially macroeconomic predictors – never completely fades as they remain important forecasting variables in the experiments that follow, but from this point onwards our attention is mostly devoted to asset returns because they are the variables for which an assessment of the economic value of using forecast combinations is most easily obtained.

5.1. *Mean Square Forecast Error Results*

When the models are flexible because of their rich parametrization, accuracy of fit is relatively unsurprising. However, rich parametrizations are also well known to introduce large amount of estimation uncertainty which may end up deteriorating their out-of-sample performance, see e.g., Rapach and Wohar (2006). In the predictability literature this has been stressed among the others by Chan, Karceski, and Lakonishok (1998) who – studying the out-of-sample predictability of stock returns – found that except for the term and default spreads, macroeconomic variables tend to perform poorly. Goyal and Welch (2003) report that the dividend yield is a good predictor of stock returns only when the forecasting horizon is longer than 5 to 10 years. Otherwise the fit is good only in-sample. Neely and Weller (2000) re-examine the findings of Bekaert and Hodrick (1992) using a predictive metrics. They show that VAR models are outperformed by much simpler benchmarks, like random walks. They suggest that the poor forecasting performance may be due to underlying structural changes. Similarly, Bossaerts and Hillion (1999) argue that even the best linear models contain no out-of-sample forecasting power even when the specification of the models is based on statistical criteria that should penalize over-fitting. They speculate that the parameters of the selected models may be changing over time so that the correct model ought to be nonlinear, possibly of a regime switching type.

To assess whether taking regimes into account leads to any improvement in prediction performance, we implement the following 'pseudo out-of-sample' recursive strategy. For a given model, we obtain recursive estimates over expanding data windows 1926:12 - 1985:01, 1926:12 - 1985:02, etc. up to 1926:12 - 2004:12. This gives a sequence of 240 sets of parameter estimates specific to each of the models. For instance, the regime switching model (3) generates 240 sets of regime-specific intercepts, covariance matrices, and transition matrices, as well as regime-independent VAR(1) coefficients. At each final date in the expanding sample – i.e. on 1985:01, 1985:02, etc. up to 2004:12-h, where h is the forecast horizon (in months) – we calculate h-month ahead forecasts for each of the 8 variables under study, i.e. including both financial and macroeconomic variables. For concreteness, in what follows we focus on $h = 1$, 12 months. Calling $\hat{y}_t^{(M,h)}$ the forecast generated at time t, by model M, at horizon h (in the following we drop the variable index, $n = 1, ..., 8$).

Finally, we evaluate the accuracy of the resulting forecasts, by calculating the resulting forecast errors, $e_t^{(M,h)} \equiv y_{t+h} - \hat{y}_t^{(M,h)}$.

Table 3 reports summary statistics concerning the quality of relative forecasting performance. In particular, we report three statistics illustrating predictive accuracy: the root mean-squared forecast error (RMSFE),

$$RMSFE^{(M,h)} \equiv \sqrt{\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left(y_{t+h} - \hat{y}_t^{(M,h)}\right)^2},$$

the prediction bias

$$Bias^{(M,h)} \equiv \frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left(y_{t+h} - \hat{y}_t^{(M,h)}\right),$$

and the standard deviation of forecast errors,

$$SD^{(M,h)} \equiv \sqrt{\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left[\left(y_{t+h} - \hat{y}_t^{(M,h)}\right) - \frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h}\left(y_{t+h} - \hat{y}_t^{(M,h)}\right)\right]^2}.$$

Notice that the three statistics are not independent as it is well known that $MSFE^{(M,h)} \equiv \left[Bias^{(M,h)}\right]^2 + \left[SD^{(M,h)}\right]^2$, i.e. the MSFE can be decomposed in the contribution of bias and forecast error variance. At a one-month horizon, the four-state model seems to be rather useful for forecasting purposes, as it displays the lowest MSFE for five out of eight variables. When compared to the other six benchmarks, the reduction in MSFE is particularly important for excess stock returns and real money growth (the ratio of the four-state MSFE to the next best MSFE are 0.88 and 0.82, respectively). However for two variables (excess bond returns and the default spread) simple random walk benchmarks outperform all other models, although the difference relative to the four-state model is small. Consistently with the results in Section 4.2, when $h = 1$ a richer three-state model with regime-dependent VAR coefficients is systematically dominated by the four-state model in terms of MSFE minimization; also, a VAR(1) seems too simple to be able to produce useful forecasts.[17] Moreover, also Pesaran-Timmerman (PT) recursively updated regressions and the univariate AR models provide disappointing performances, especially for what concerns excess asset returns and real money growth. Finally, the random walk benchmarks produce biases that are close (or even lower) to those characterizing the four-state model, the difference being in the variances of the forecast errors: regime switching models produce higher biases (i.e. they often miss the actual point value of the forecast variable), but they seem to be able to systematically move in a way that reduces the error variance. This is what one expects if the model's regimes well identify in real time the economy's turning points.

Table 3 approximately here

At longer horizons, the superior performance of the four-state model tends to deteriorate slightly, while the three-state model acquires merit. For instance, at $h = 12$ the MSIAH(3,1) presents the lowest MSFE for half of the variables, although the four-state model still outperforms all benchmarks in predicting excess stock returns, the dividend yield, and real money growth. The performance of PT-style recursive regressions remains slightly inferior to either regime switching (e.g., excess returns and inflation), or falls very close to the accuracy

[17] The exceptions are the default spread, which is best modeled as a three-state process, and the inflation rate which is best predicted by a simple VAR(1).

of simpler models (VAR and random walks). Interestingly, the family of regime switching models comes to consistently outperform single-state (as well as naive) benchmarks at longer horizon: this is the case for five out of eight variables at $h = 12$. This is relatively unsurprising in the light of the evidence we have offered in Section 4 of the fact that regime switching models generally perform well at fitting the multivariate density of the data: as h grows, forecasts from models with regimes rely less and less on the accurate identification of the most recent turning points of the economy, and increasingly on the overall properties of the unconditional density of the data.

5.2. *Testing Differential Predictive Accuracy*

Further examination of Table 3 reveals that – although rankings match our general comments – differences between models are often modest, which casts some doubts on how useful the four-state model may be in practice. We therefore proceed to assess whether the out-of-sample performances are different enough to allow us to draw any conclusions on the relative precisions of alternative models by implementing the forecast accuracy comparison tests proposed by Diebold and Mariano (1995). For concreteness, we focus on the forecasts of excess stock and bond returns only. Let

$$dif_t^{(m,n,h)} \equiv \left(e_t^{(m,h)}\right)^2 - \left(e_t^{(n,h)}\right)^2$$

be the differential loss (in our case the standard square loss) of model m relative to the loss from model n. We can test the significance of the differences between two sets of forecast errors based on the statistic

$$DM_h^{(m,n)} = \frac{\frac{1}{240-h}\sum_{t=1985:01}^{2004:12-h} dif_t^{(m,n,h)}}{\widehat{\sigma}(dif_t^{(m,n,h)})} \overset{a}{\sim} N(0,1), \tag{9}$$

where $\widehat{\sigma}(dif_{t,h}^{(m,n)})$ is an estimate of the standard error of the loss function differential, e.g. the HAC estimator

$$\widehat{\sigma}(dif_{t,h}^{(m,n)}) = \sum_{j=-h}^{h} Cov\left[dif_{t,h}^{(m,n)}, dif_{t+j,h}^{(m,n)}\right].$$

Tables 4 and 5 illustrate the results of such tests. Table 4 concerns $h = 1$ and Table 5 $h = 12$. Moreover, values of $DM_h^{(m,n)}$ appearing below the main diagonal concern forecasts of excess stock returns, while those above the main diagonal refer to forecasts of excess bond returns. A negative value of $DM_h^{(m,n)}$ implies that the model in row (m) outperforms the model in column (n) as it produces a lower sample MSFE loss. Finally, in the tables we boldface values of $DM_h^{(m,n)}$ which imply p-values of 0.05 or lower under Diebold and Mariano's (1995) standard asymptotic normal distribution. However – as discussed by West (1996) – caution should be exercised when interpreting these results because the small sample distribution of the statistic in (9) can show large departures from normality, thus complicating the task of conducting inferences on differential predictive accuracy. We tackle the issues posed by possible departures of the small-sample distribution of (9) from normality by block-bootstrapping the distribution of $DM_h^{(m,n)}$ for all pairs of models and $h = 1, 12$ months. The simulation procedure consists of three steps:

1. Generate B bootstrap resamples of the sample indices $\{1, 2, ..., T\}$ where T is the length of the out-of-sample period. The block length L is set equal to the forecast horizon, h. This choice is justified by the

overlapping nature of the realizations underlying the forecast error series.[18] The bootstrap resampling under a block length of L can be easily performed as follows:

(a) Determine the number of blocks v needed to span the entire sample size T, i.e. $v \equiv int(T/L) + 1$ where $int(\cdot)$ denotes the integer part of the rational number T/L.

(b) For $a = 1, ..., v$ draw τ_a^b as a discrete IID uniform over $\{1, 2, ..., T\}$ obtaining a collection of v time indices $\{\tau_1^b, \tau_2^b, ..., \tau_v^b\}$

(c) The b-th resample indices are then given as $\{\tau_1^b, \tau_1^b + 1, ..., \tau_1^b + L - 1, \tau_2^b, ..., \tau_2^b + L - 1, ..., \tau_v^b, ..., \tau_v^b + L - 1\}$. If it occurs that for some $j \geq 1$ $\tau_a^b + j > T$, then set the time index to $\tilde{\tau}_a^b = \tau_a^b + j - T$, i.e. the resampling is restarted from the beginning of the sample in case the scheme attempts to draw indices that exceed the overall sample size T.

(d) repeat a.-c. for $b = 1, 2, ..., B$.

2. For each of the B resamples of length T indexed by b we calculate (9), $\widehat{DM}_h^{(m,n)}$.

3. At this point a small-sample simulated distribution of $DM_h^{(m,n)}$ is obtained , $\{\widehat{DM}_{h,b}^{(m,n)}\}_{b=1}^B$. We report p-values for two-sided tests of the null of zero differential predictive accuracy, i.e. the percentage of the bootstrapped simulations such that

$$|\widehat{DM}_{h,b}^{(m,n)}| > |DM_h^{(m,n)}|,$$

where $DM_h^{(m,n)}$ corresponds to the sample value of the statistic.

Table 4 approximately here

In the bootstrap implementation in the paper we set $B = 50,000$ while the overall sample size T clearly depends on the forecast horizon h. Tables 4 and 5 report the bootstrapped p-values in parenthesis, below the values for (9). Table 4 shows that over short horizons the MSIH(4,0)-VAR(1) provides exceptional performance for excess stock returns, since the model outperforms all other forecasts in a statistical significant fashion. The associated p-values are in fact below 0.01 under the asymptotic distribution for (9), while they generally fall around 0.05 when the small sample distribution of the statistic is bootstrapped. The only exception concerns the comparison to the random walk, where the asymptotic p-value exceeds 0.05 and the bootstrapped one is relatively large (0.16). Interestingly, we uncover evidence that the richer three-state MSIAH(3,1) model outperforms a few of the other forecasts, although in this case the bootstrapped p-values tend to fall around 0.1. When it comes to excess bond returns, Table 4 shows that only two conclusions can be reached. First, the four-state model outperforms in a significant way PT's predictive regressions and the AR models. Second, in this case the random walk guarantees an interesting precision, for example significantly outperforming PT's methods and the three-state model.

Table 5 approximately here

[18] For instance, a negative return for some of the assets at time t that falls in the extreme left tail of the predictive density under some given statistical model is likely to generate large and negative forecast errors from all 12-month ahead forecasts generated between time $t - 12$ and $t - 1$. To avoid considering these sub-sequences of errors as an indication of structure in the forecast error series, we set the block length to h.

Table 5 presents results concerning 12-month forecasts. Importantly, results are qualitatively similar – only marginally weaker in terms of bootstrapped p-values – than in Table 4. The four-state model keeps outperforming most other competitors with some statistical strength when it comes to predict excess stock returns. However the MSIAH(3,1) framework is now inferior to all other forecasts, which is consistent with the idea that a three-state model might be not correctly capturing the shape of the ergodic density of returns data. Interestingly, the performance of the four-state model improves (vs. $h = 1$) in terms of performance on excess bond returns, since now the implied MSFE is significantly lower to the one produced by the three-state, the VAR(1) and the random walk. The performance of the single-state VAR(1) model is particularly poor.

In summary, predictive accuracy tests show that differences in Table 3 could not entirely attributed to chance. Consistently with much forecasting literature, at short horizons it remains true that outperforming the random walk or simple constant expected returns benchmarks is difficult. At longer horizon, it is clear that the flexibility of mixture (regime switching) models may be required to capture the salient features of the multivariate distribution of financial and macroeconomic variables. However, careful scrutiny of Tables 4 and 5 also highlights that predictive accuracy comparisons are hardly unidirectional. We find a few instances in which while model A fails to significantly outperform model B, and B fails to outperform C, the same cannot be said about A outperforming C.[19] We take this as an indication that relative performances are rather heterogeneous on different parts of the (pseudo) out-of-sample period. This opens the door to the possibility that carefully combining forecasts produced by different frameworks may deliver important payoffs. This is the issue that the next two Sections explore.

6. **Are Forecast Combination Useful? Statistical Evidence**

The simultaneous availability of forecasts from linear and non-linear, from univariate and multivariate, and from models with constant vs. time-varying parameters represent one of the classical cases in which forecast combinations are generally thought of as valuable: unless a particular forecasting model that (uniformly, i.e. for each out-of-sample period) generates smaller forecast errors than all of its competitors can be identified, forecasts combinations will always ex-ante reduce the overall MSFE because they offer "diversification gains". The practical value of this insight for our case study is analyzed in this Section.

6.1. *The Forecast Combination Problem*

We are interested in forecasting at time t the future values after $h \geq 1$ months of the first two elements of $\mathbf{y}_t$, i.e. excess stock and bond returns, $\mathbf{x}_{t+h} \equiv [x^s_{t+h}\ x^b_{t+h}]'$. Call $\hat{\mathbf{x}}^s_{t,t+h}$ and $\hat{\mathbf{x}}^b_{t,t+h}$ two $F \times 1$ vectors – one for excess stock returns, the other for excess bond returns – that collect F distinct forecasts produced at time t. Finally, call $\tilde{x}^s_{t,t+h}$ and $\tilde{x}^b_{t,t+h}$ the forecast combinations obtained as

$$\begin{aligned}\tilde{x}^s_{t,t+h} &= g^s(\hat{\mathbf{x}}^s_{t,t+h}; \boldsymbol{\omega}^s_{t,t+h}) \\ \tilde{x}^s_{t,t+h} &= g^b(\hat{\mathbf{x}}^b_{t,t+h}; \boldsymbol{\omega}^b_{t,t+h}),\end{aligned}$$

[19] For instance, for $h = 12$ and with reference to excess bond returns (panel above the main diagonal), both MSIH(4,0)-VAR(1) (p-value is 0.15) and the random walk fail to outperform the AR model (p-value 0.10), suggesting that MSIH(4,0)-VAR(1) and the random walk must be providing "similar" performances. However MSIH(4,0)-VAR(1) significantly reduces the MSFE vs. the random walk (p-value is 0.01).

where $g^{s}(\cdot;\boldsymbol{\omega}^{s}_{t,t+h})$ and $g^{b}(\cdot;\boldsymbol{\omega}^{b}_{t,t+h})$ are functions ($\mathcal{R}^{F}\rightarrow\mathcal{R}$) – parameterized by two $(F+1)\times 1$ vector of weights $\boldsymbol{\omega}^{s}_{t,t+h}$ and $\boldsymbol{\omega}^{b}_{t,t+h}$ – that *combine* (pool) the F forecasts into one aggregate forecast. For instance, $g^{s}(\cdot;\boldsymbol{\omega}^{s}_{t,t+h})$ and $g^{b}(\cdot;\boldsymbol{\omega}^{b}_{t,t+h})$ might be restricted to be linear affine functions, in which case

$$g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};\boldsymbol{\omega}^{[\cdot]}_{t,t+h})=\omega_{0,t,t+h}+\sum_{f=1}^{F}\omega_{f,t,t+h}\hat{x}^{[\cdot]}_{t,t+h},\tag{10}$$

where the subscript $[\cdot]$ refers to either stocks or bonds.[20] For instance, when an equal weighting scheme is imposed, then $\omega_{f,t,t+h}=1/F$, $f=1,...,F$ and $\omega_{0,t,t+h}=0$.

In general, while the structure of the functions $g^{[\cdot]}(\cdot;\boldsymbol{\omega}^{s}_{t,t+h})$ tends to imposed by the researcher, the combination weights are not exogenously fixed, but they are instead computed using the principle that decision makers do not attach any intrinsic value on forecasts: it is instead the forecasts (or combination thereof) that minimize their overall loss function to matter. In particular, we follow standard practice and assume that the loss function only depends on the forecast error, $e^{s}_{t,t+h}\equiv x^{s}_{t,t+h}-\tilde{x}^{s}_{t,t+h}$, $e^{b}_{t,t+h}\equiv x^{b}_{t,t+h}-\tilde{x}^{b}_{t,t+h}$, and that the loss function is of a standard, symmetric quadratic type:

$$L\left(e^{[\cdot]}_{t,t+h}\right)\equiv\left(e^{[\cdot]}_{t,t+h}\right)^{2}.$$

Therefore the forecast combination problem becomes

$$\min_{\boldsymbol{\omega}^{[\cdot]}_{t,t+h}} E\left[\left\{x^{[\cdot]}_{t,t+h}-g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};\boldsymbol{\omega}^{[\cdot]}_{t,t+h})\right\}^{2}\Big|\,\hat{\mathbf{x}}^{[\cdot]}_{t,t+h}\right],\tag{11}$$

i.e. the forecast combination weights are chosen to minimize the (conditional) MSFE. In this case, it is straightforward to show that

$$\tilde{\boldsymbol{\omega}}^{s}_{t,t+h}\ \ni g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};\tilde{\boldsymbol{\omega}}^{s}_{t,t+h})=E\left[x^{[\cdot]}_{t,t+h}\Big|\,\hat{\mathbf{x}}^{[\cdot]}_{t,t+h}\right],\tag{12}$$

i.e. the weights will be chosen to allow $g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};\tilde{\boldsymbol{\omega}}^{s}_{t,t+h})$ to best approximate the conditional expectation of the excess asset returns given the information contained in the $F\times 1$ vector of forecasts. For concreteness, and following in the steps of Bates and Granger (1969), we will restrict our attention to the case in which $g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};\boldsymbol{\omega}^{[\cdot]}_{t,t+h})$ is a linear function, as in (10). This choice together with (12) implies that $\tilde{\boldsymbol{\omega}}^{s}_{t,t+h}$ can be simply estimated as a vector of least squares estimates in the regression:

$$x^{[\cdot]}_{t,t+h}=\omega_{0,t,t+h}+\sum_{f=1}^{F}\omega_{f,t,t+h}\hat{x}^{[\cdot]}_{t,t+h}+\eta^{[\cdot]}_{t,t+h}.\tag{13}$$

As shown by Chong and Hendry (1986), under squared loss function (13) also provides a test of pair-wise encompassing. For instance, when $F=2$ $\omega_{0,t+h}=\omega_{2,t+h}=0$, $\omega_{1,t+h}=1$ implies that forecast 1 encompasses forecast 2, while $\omega_{0,t+h}=\omega_{1,t+h}=0$, $\omega_{2,t+h}=1$ brings to the conclusion that forecast 2 encompasses 1. When one forecast function encompasses another, this means that combining the two forecasts will not reduce the overall MSFE. However, an indication that for some of the forecasts in (13) the corresponding weight is nil, will automatically lead to avoid using the forecast combination, in the sense that (trivially) $g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};[0\ 1\ 0]\prime)=\hat{x}^{1}_{t,t+h}$ and $g^{[\cdot]}(\hat{\mathbf{x}}^{[\cdot]}_{t,t+h};[0\ 0\ 1]\prime)=\hat{x}^{2}_{t,t+h}$.

[20] Granger and Ramanathan (1984) recommend adding a constant intercept under squared loss; Elliott and Timmermann (2004) reach the same conclusion for a variety of loss functions.

Because

$$E\left[\left\{x_{t,t+h}^{[\cdot]}-\omega_{0,t,t+h}-\sum_{f=1}^{F}\omega_{f,t,t+h}\hat{x}_{t,t+h}^{[\cdot]}\right\}^{2}\middle|\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}\right]=Var\left[\tilde{e}_{t,t+h}^{[\cdot]}\middle|\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}\right]+\left\{E\left[\tilde{e}_{t,t+h}^{[\cdot]}\middle|\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}\right]\right\}^{2},$$

($\tilde{e}_{t,t+h}^{[\cdot]}$ is defined as the forecast combination error) problem (11) implies that each of forecast functions entering $\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}$ is useful in the measure in which they help predicting either the (conditional) bias of the pooled forecast, or its conditional variance. As stressed by Timmermann (2005), it is common to find that $E[\tilde{e}_{t,t+h}^{[\cdot]}|\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}]$ exceeds the bias produced by the best forecasting model in the vector $\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}$; however, in practice forecast combinations often reduce the MSFE by reducing $Var[\tilde{e}_{t,t+h}^{[\cdot]}|\hat{\mathbf{x}}_{t,t+h}^{[\cdot]}]$ by substantial amounts.

Besides equal-weighting schemes – i.e. $\omega_{f,t,t+h}=1/F$, $f=1,...,F$ and $\omega_{0,t,t+h}=0$ – another popular combination benchmark suggested by Winkler and Makridakis (1983) and Stock and Watson (2001) consists of computing weights that simply reflect the performance of each individual model relative to the performance of the average model but ignore correlations across forecasts. As first remarked by Newbold and Granger (1974), in small samples and with large numbers of forecast functions F, correlation may prove hard to estimate with any precision. Letting $MSFE_{f,t,t+h}\equiv\sum_{\tau=1}^{t}(x_{\tau}-\hat{x}_{f,\tau,\tau+h})^2$ the mean-squared error from model $f=1,...,F$, Stock and Watson suggest to set

$$\omega_{f,t,t+h}=\frac{[1/(MSFE_{f,t,t+h})^{\alpha}]}{\sum_{f=1}^{F}[1/(MSFE_{f,t,t+h})^{\alpha}]} \tag{14}$$

and find that the scheme works well in practice. In the following we use equation (14) setting $\alpha=1$, i.e. we experiment with a scheme that assign the higher weight the lower is the MSFE.

Finally, an important issue often debated in the forecast combination literature is whether any constraints ought to be imposed on the weights $\boldsymbol{\omega}_{t,t+h}$, chiefly whether it makes sense to either impose that $\boldsymbol{\omega}_{t,t+h}'\boldsymbol{\iota}_F=1$ (i.e. the combination weights must sum to one) or that $0\leq\omega_{f,t,t+h}\leq 1$ $f=1,...,F$ (i.e. to rule out that the combined forecasts may lie outside the range of individual forecasts). On one hand, rigorous arguments have been made in favor of both restrictions: when the F forecasts are all unbiased, $\boldsymbol{\omega}_{t,t+h}'\boldsymbol{\iota}_F=1$ guarantees that the forecast combination will also be unbiased; under the assumption of unbiasedness, Diebold (1988) argues that $\boldsymbol{\omega}_{t,t+h}'\boldsymbol{\iota}_F=1$ minimizes the chances that the forecast errors from the combination may be serially correlated and hence predictable (viz. inefficient). The convexity restriction $0\leq\omega_{f,t,t+h}\leq 1$ avoids the possibility that – given a range of possible $h-$step ahead forecasts at time t – the combined forecast may lie outside the scope of the originally predicted values. On the other hand, the empirical results to follow will offer plenty of evidence of the presence of biases in the forecasts. This is rather intuitive when facing structural shifts in the intercept of the process followed by the variables of interested, as documented in Section 4.2. Additionally, and similarly to Bunn (1985) and Guidolin and Timmermann (2005b), it is important to recognize that the fact that a combination weight is negative does not imply it has no value and negative combination weights can and do help providing more precise forecasts than the original input forecast functions.[21]

[21]Tests of both restrictions led to strong rejections for all the least-squares combination regressions that follow.

6.2. *Recursive Estimates of Combination Weights*

We start by computing recursive combination weights over our (pseudo) out-of-sample period, i.e. over sequences of expanding windows of data 1985:01 - 1985:12, 1985:01 - 1986:01, etc. up to 1985:01 - 2004:12$-h$.[22] For a variety of model combinations, we do that in two ways:

- By estimating the weights from regressions of type (13) and in which F is set to either two (i.e. combinations for pairs of models) or to seven, which is the total number of forecasting models introduced in Section 2. In the following we call these weights "OLS weights".
- By setting the weights according to equation (14), when $\alpha = 1$ and $F = 7$. In the following we call these weights "inverse-MSFE weights".

As anticipated, we also compute equally weighted forecast combinations with $F = 7$. It must be noticed that pairwise OLS weights are also computed for the case in which two different regime switching forecasts – the ones from the MSIH(4,0)-VAR(1) and those from the MSIAH(3,1) – are combined. This is a very interesting experiment that might provide additional prediction accuracy in the case in which three- and four-state models pick up structural shifts at different frequencies or reflecting heterogeneous economic forces. Using the combination weights obtained in these three alternative ways, we proceed to quantify predictive performance along lines similar to Section 5.

Figure 5 provides a few examples of the behavior of the recursively estimated weights. The first two plots in panel A concern OLS weights for one specific pairwise combination, the one between the four-state model and PT-type forecasts, when the predictors are selected by recursive maximization of the $\bar{R}^2$.[23] Especially in the case of excess stock returns, Figure 5 displays enormous variability in the estimated combination weights. In particular, while the weight assigned to the regime switching forecasts hovers in the narrow range 0.55-0.58, the weight received by the PT forecasts is highly variable, oscillating from very high values in excess of 14 to the negative range. Even the intercept correction, although seemingly negligible in the scale of the plots, jiggles a lot in the range -0.48 to 0.02. Clearly, such wild gyrations in combination weights produce large amounts of parameter uncertainty that are likely to degrade the out-of sample performance of the combinations.

Figure 5 approximately here

Panel B of Figure 5 shows instead the inverse-MSFE weights for both excess stock and bond returns. Consistently with our comments to Table 3, in the case of excess stock returns the four state model systematically receives a large weight (roughly 24%), followed by the three state model (around 15%), which however gradually declines over time. The remaining 60% is allocated in approximately equal fashion across the remaining five models, with the sigle-state VAR(1) receiving a slightly lower weight of 10-11%. In the case of excess bond returns, weights are much more balanced and stable over time, with approximately 17% going to the VAR(1) and the random walk, 15-16% each to the two regime switching forecasts, and the remaining third to PT and AR univariate forecast functions. Clearly, inverse-MSFE weights imply substantially lower parameter uncertainty.

[22] To avoid excess instability in the coefficients in the first part of the recursive experiment, we use the first 12 values of forecast functions and prediction errors to provide the 12 initial values of the optimal weight series (not shown in the figures).

[23] Additional plots concerning other pairwise combinations or 12-step ahead combination weights appear qualitatively similar and are not reported to save space. They are available upon request.

6.3. *Statistical Tests*

Tables 6 and 7 provide information on the statistical performance of forecast combinations for $h = 1$ and 12 months, respectively. In both tables, panels A1 show the results for pairwise forecast combinations "centered" around the MSIH(4,0)-VAR(1) model, i.e. in which $f = 1$ is held fixed at the four-state forecast combinations. Panels B1 report on similar experiments when the combinations are "centered" around the PT forecasts when the predictors are recursively selected by maximization of the $\bar{R}^2$. We adopt this additional benchmark as a way of a robustness check; we select this PT-style model because of its slightly better performance (for $h = 1$) vs. similar univariate prediction models in Table 3. Finally, in both Tables 6 and 7, panels A2 and B2 show results for a few benchmarks which do not imply selection of forecasts to be combined: when all forecasts are combined by computing OLS weights ($F = 7$); equally-weighted forecast; inverse-MSFE weights. Additionally, for reference, we copy the performance statistics for our favorite regime switching framework, the four-state model.

For the 1-month horizon, Table 6 shows that forecast combinations "centered" around the four-state model fail to improve the MSFE. On the opposite, they increase the MSFE vs. just using the MSIH(4,0)-VAR(1) model, from 3.8% to approximately 6.1% for stocks, and from 1.3% to 2.8% in the case of bonds. Interestingly, most of the deterioration is attributable to an increase in the variance of the forecast errors. This is likely due to the enormous variability in the estimated combination weights and in the resulting large parameter uncertainty, see Figure 5. Panel B shows similar results for OLS weights centered on PT forecasts. In fact, in this case there is even evidence that combining PT forecasts with regime-switching based predictions delivers results which are sensibly inferior to all other sorts of combinations. Results hardly improve when F stops being forced to be equal to two and all of the forecast functions are part of the combination.

Table 6 approximately here

The fact that forecast combinations may be susceptible to failure in the presence of large estimation errors in the combination weights has been recently discussed by Yang (2004). Finite-sample errors in the estimates of the combination weights can lead to poor performance of combination schemes that can be proven to dominate in large samples, in part because of the prevalence of strong multicollinearity among the forecasts that are being combined. In the same way in which non-stationarities and structural change may provide a justification for the use of forecast combinations, they can also lead to instabilities in the combination weights and deteriorate the out-of-sample performance of the combination, see e.g., Palm and Zellner (1992).

However panels A2 and B2 of Table 6 show that when parameter uncertainty related to estimation of the combination weights is wiped out – by using an equal weighting scheme – pooling forecasts delivers interesting improvements in forecast accuracy. The RMSFE declines to 2.2% for excess stock returns and to 1.1% in the case of excess bond returns; both values are lower than the best among pairwise combinations and the four-state model. The improvement is almost entirely explained by a reduction in the forecast error variance, which is again consistent with the idea that in panels A1 and B1 the high variability of the OLS weights is what causes the failure of pairwise pooling. This result echoes previous studies which have found that the simple weighting schemes are generally hard to beat. The reasons of this fact are discussed in Dunis, Laws, and Chauvin (2001) and Hendry and Clements (2002). The most convincing explanation is that the estimation uncertainty of the optimal weights offsets the advantage of using the forecast combination.

When estimation uncertainty is "intelligently" dealt with – by setting the combination weights to be relative performance indicators – the improvement is substantial and across the board, i.e. it involves both excess stock and bond returns. Under inverse-MSFE weights, the RMSFE error declines to 1.6% for excess stock returns and to 0.7% for bonds. These values are relatively small and certainly dwarfed by the magnitude of the volatility of both series (5.5 and 1.9 percent, respectively).

Table 6 shows that in the presence of structural change/regimes, forecast combinations improve prediction accuracy, although this result may be severely limited by the presence of pervasive estimation uncertainty when the weights are estimated, for instance using regression schemes a' la Granger and Bates (1969). In fact, the MSFE improvement from using inverse-MSFE weights in forecast combinations turns out to be statistically significant: the corresponding Diebold-Mariano (DM) test for differences in MSFEs between the pooled prediction and the four-state model produces a p-value of essentially 0.00 (the DM-stat is 7.37) for excess stock returns and again of 0.00 (the DM stat is 9.34) for excess bond returns.[24] Similarly, when the baseline model is the PT scheme, the implied p-values for a test of the null that equal-weights and inverse-MSFE weights combinations predict as well as PT are around 0.01, with DM-stats of around 5 for excess stock returns and of 10 for excess bond returns (in the latter case the p-value is essentially nil).

Finally, Table 7 extends these tests to a 12-month forecast horizon. Panels A1 and B1 show that the results in Table 6 also apply to 1-year forecasts, although in this case the improvement brought about by the combinations is more modest, e.g. from 4.6% for the four-state model to 4.1% in the case of stocks, and from 2.1% to 1.9% in the case of bonds. Once more, inverse-MSFE combinations turn out to be the best, while combinations (even relatively tightly parameterized ones, with $F = 2$) based on OLS weights are not very successful. However such improvements are no longer statistical significant. When the baseline model is the four-state regime switching (panel A1), even the inverse-MSFE weights imply a p-value of 0.23 (the DM stat is 1.81) for stocks and a p-value of 0.14 (the DM stat is 1.58) for bonds. In conclusion, forecast combinations work well with short investment horizons and when estimation of the weights can be somehow constrained or even avoided altogether.

Table 7 approximately here

7. The Economic Value of Forecast Combinations: Portfolio Implications

The finding that forecast combinations may lower the RMSFE for excess stock and bond returns fails to pin down the more interesting issue of their actual economic value to decision makers. In fact, although many decision makers might have an interest for the results so far and the general issue of whether the dynamic linkages between financial markets and macroeconomic factors might have been subject to regime shifts, there is a class of economic agents that has a straightforward use for the results in Tables 6 and 7: portfolio managers. In other words, the squared loss function we have adopted so far has a lot of convenience to it, but need not represent a realistic loss to decision makers. It seems that portfolio returns, their Sharpe ratio, or some utility function of realized wealth may represent a much more informative criterion to decide whether forecast combinations have any value.

In fact, money managers have a keen interest in exploiting statistical predictability to improve the return-risk properties of their portfolios. In this sense, such decision makers would be mostly interested not in point

[24] The same applies to equally weighted forecasts: the p-value for a test comparing to the four-state model is 0.00 for both excess stock and bond returns, with DM stats of 7.12 and 9.86. These p-values and the ones mentioned in the rest of the Section were bootstrapped using the same algorithm discussed in Section 5.2.

forecasts of future asset returns or in the ability of (3) to approximate their conditional joint density, but in correctly forecasting (one-step ahead) Sharpe ratios,

$$SR_{t,t+1}^{(M,i)} \equiv \frac{E_t^{(M)}[y_{t+1}^i]}{\sqrt{Var_t^{(M)}[y_{t+1}^i]}},$$

where i indexes excess stock and bond returns, respectively.

Figure 6 shows 1-month ahead, predicted Sharpe ratios for both stocks and bonds. Such predictions are calculated under each of the seven models introduced in Section 2. As in Sections 5 and 6, the ratios are computed recursively, by updating parameter estimates over the expanding windows 1926:12 - 1985:01, 1926:12 - 1985:02, etc. up to 1926:12 - 2004:12.[25] Although most models imply wild variation in the one-month ahead prediction of the Sharpe ratios, a few qualitative facts can be noticed. First, as one would expect, models with regimes imply wide swings in $SR_{t,t+1}$, although a structural difference seems to exist between MSIH(4,0)-VAR(1) and MSIAH(3,1), in the sense that the latter model implies sharper corrections and a rather implausible tendency for $SR_{t,t+1}$ to be predicted in the negative range. In fact, while the four-state model implies a sensible sample mean of 0.10 for the equity Sharpe ratio (0.02 for bonds), the same cannot be said for the three-state model, where the sample mean is negative (-0.02, the same number obtains for bonds), which is inconsistent with equilibrium asset pricing principles.[26] Second, as already noticed in a related experiment by Guidolin and Ono (2006), the single-state model produces Sharpe ratios that are systematically too low for both asset classes, and systematically lower for stocks than for bonds, which is also counter-intuitive. Third, regression-based models predict values of $SR_{t,t+1}$ for equities which are essentially constant over time and that seem to fall at average levels which depart from the historical values commonly used in finance, e.g. in excess of 0.25 per month vs. a historical level of approximately 0.11. Notice that flat predictions for the Sharpe ratio are hardly compatible with active portfolio management. Fourth, the four-state model gives reasonable risk-return insights. The equity Sharpe ratio fluctuates over time in a counter-cyclical manner, i.e. $SR_{t,t+1}^{(M,s)}$ is high during recessions and declines during economic booms. Under regime switching, the bond $SR_{t,t+1}^{(M,b)}$ is stable but still provides strong and possibly useful signals.

Figure 6 approximately here

7.1. *Recursive Portfolio Weights*

In this section we proceed to the systematic, recursive calculation of optimal mean-variance portfolio weights and assess the comparative out-of-sample portfolio performance of (the best performing) forecast combinations vs. the models introduced in Section 2. Assume an investor endowed with initial unit wealth has preferences described by a simple mean-variance functional:

$$\begin{aligned} V_t &= E_t[W_{t+1}] - \frac{1}{2}\lambda Var_t[W_{t+1}] \\ W_{t+1} &= w_t^s(1 + x_{t+1}^s + r_t) + w_t^b(1 + x_{t+1}^b + r_t) + (1 - w_t^s - w_t^b)(1 + r_t), \end{aligned} \tag{15}$$

[25] Notice the existence of a structural difference between the regime switching and the remaining models, because while the former explicitly imply time-variation in variances as well as risk premia, the regression-based (e.g. PT schemes, and univariate AR) and random walk models allow variation in variances only as a result of the recursive updating of the estimates.

[26] Interestingly, the four-state model produces sample means for the Sharpe ratios which are virtually identical to the random walk model (no predictability in risk premia), despite the variability in the ratios themselves.

where λ is interpreted as coefficient of risk aversion that trades-off (conditional) predicted mean and variance of the one-step ahead wealth, and r_t is a short-term (1-month) interest rate.[27] At each time t in the sample, the investor maximizes V_t by selecting weights $\mathbf{w}_t \equiv [w_t^s \ w_t^b]'$ when the predicted moments are calculated using some reference statistical model, e.g. our four-state model or some forecast combination scheme. Simple algebra shows that:

$$\hat{\mathbf{w}}_t = \frac{1}{\lambda}\{\widehat{Var}_t[\mathbf{x}_{t+1}]\}^{-1}\{\hat{E}_t[\mathbf{x}_{t+1}]\},$$

Portfolio weights are calculated recursively using the recursive parameter estimates also employed in Sections 5 and 6.[28] Problem (15) is solved when investors are prevented from selling any securities short, i.e. such that $\mathbf{w}_t'\mathbf{e}_n \in [0,1]$ $\forall t$ and $n = s,\, b$.[29] Of course, (15) corresponds to a very simplistic asset allocation strategy in which preferences (absolute risk aversion) are assumed to be constant, and investors care only about mean and variance. Additionally, such investors would have to be strictly myopic and ignore time-varying investment opportunities. Notice that regime switching models imply rich (as well as dynamic, as the state probabilities are recursively updated) departures from standard zero skewness and constant kurtosis levels, which may be at odds with the assumption of mean-variance preferences, especially for horizons exceeding one month. However, this seems to be a relatively straightforward way for us to collect some evidence bearing on the issue of whether forecast combinations have any economic value for a portfolio manager.[30]

Table 8 shows a few statistics for recursive optimal portfolio weights obtained under four alternative assumptions on the risk aversion coefficient, $\lambda = 0.2$, 0.5, 1, and 2.[31] The table shows that different models/forecast combination schemes imply quite different average allocations to the different assets, as well as different variability of the optimal portfolio weights. For instance, assuming $\lambda = 1$, the percentage invested in stocks ranges from 100 under the random walk or PT-style predictions (both implying high predictions of the equity Sharpe ratios) to 17 under a VAR(1) model; the weight assigned to bonds similarly ranges from 0 to 43 percent. Interestingly, regime switching models and forecast combination schemes end up implying comparable portfolio weights, of approximately 60% in stocks, 20% in bonds, and 20% in cash. Notice that this result is far from trivial because the combination schemes used in Table 8 involve with non-negligible weights *all* the prediction models, as shown in Section 6.1. Finally, regime switching-based portfolio strategies imply a structurally higher volatility of the optimal portfolio weights. At this point the key question becomes: which portfolio allocation strategy pays off the most over a long sample such as ours?

Table 8 approximately here

[27] Notice that predictions of the interest rate fail to enter the portfolio problem because the (nominal) yield of 1-month T-bills is known ex-ante. However, in our experiments we recursively update r_t which therefore preserves a time index.

[28] For instance, $\hat{\mathbf{w}}_{1985:01}$ is based on estimated parameters obtained using data for the interval 1926:01 - 1985:01, $\hat{\mathbf{w}}_{1985:02}$ on estimates for 1926:02 - 1985:02, etc.

[29] When short-sales are restricted, $\hat{\mathbf{w}}_t$ has no closed-form solution and is calculated numerically (by grid search).

[30] Ang and Bekaert (2002a) and Guidolin and Timmermann (2005a, 2006a) explore the optimal asset allocation implications of regimes and predictability in the joint (and time-varying) distribution of excess asset returns.

[31] While $\lambda = 0.2$ is admittedly a rather low risk aversion coefficient, notice that introspection suggests that $\lambda = 2$ represents a rather high level already. For instance, it is well known that in simple mean-variance asset allocation frameworks, the optimal weight of a single risky asset is $w_{t,t+1} = SR_{t,t+1}/\lambda$.At this point, the Sharpe ratio on the value-weighted CRSP portfolio over the period 1926-2004 is 0.12; $\lambda = 2$ implies then a weight of $0.12/2 = 0.06$, i.e. only 6% in US stocks. That represents a rather conservative portfolio position. In fact the range for λ commonly spanned in the literature (see e.g. Jorion, 1986) is 0 - 2.

7.2. *Out-of-Sample Portfolio Performance*

Table 9 completes our "economic" analysis by computing (pseudo) out-of-sample, one month portfolio performance under different levels of λ and for the seven competing models/allocation strategies. Notice that models 6 and 7 now identify forecast combination-driven strategies, based on either equal forecast weights or on the inverse-MSFE weights introduced in Section 6. Additionally, we consider two benchmarks which (albeit rather naive, as they ignore all sample information on the dynamics of asset returns) are often used in practice: an equally weighted portfolio ("$1/N$", see De Miguel et al., 2006) in which 1/3 is invested in each of the available assets at each point in time; a similar "50-50" portfolio in which 50 percent shares are allocated to stocks and bonds, respectively, at each point in time. The logic of introducing these two further benchmarks is that they may end up yielding interesting performances for two reasons: First, the benchmarks are completed unaffected by parameter uncertainty, since they fail to rely on any econometric framework implying estimation issues. Second (see Section 7.4 for further details), when held over a one-month period only, these portfolios imply no need for rebalancing and thefore completely escape the incidence of transaction costs.[32]

Table 9 approximately here

In particular, we report means and medians of one-month (gross) portfolio returns, the lower and upper values of a standard 90% confidence interval (that reflects the volatility of realized portfolio performance over 1985:01 - 2004:11), and the implied Sharpe ratio that adjusts mean returns to account for risk. The last column of Table 9 also lists the realized, recursive mean-variance statistic for each model/strategy, i.e.

$$\frac{1}{239}\sum_{t=1985:01}^{2004:11}\left(\frac{W_{t+1}^{(m)}}{W_t^{(m)}}\right)-\frac{1}{2}\lambda\frac{1}{239}\sum_{t=1985:01}^{2004:11}\left\{\left[\left(\frac{W_{t+1}^{(m)}}{W_t^{(m)}}\right)-\sum_{t=1985:01}^{2004:11}\left(\frac{W_{t+1}^{(m)}}{W_t^{(m)}}\right)\right]\right\}^2, \tag{16}$$

where W_{t+1}/W_t is simply the realized gross portfolio return over a one-month holding period when model m is employed. The table reports in bold the maximum values of mean and median portfolio returns, of the Sharpe ratio, as well as of the mean-variance statistic across models.[33]

Table 9 reveals that forecast combinations outperform all other models in all cases and under all dimensions – mean and median portfolio returns, Sharpe ratio, and a mean-variance objective. In particular, the performance indicators are systematically best for the inverse-MSFE weighting scheme of Stock and Watson (2003), although also equal weights tend to perform well. For instance, assuming $\lambda = 1$, inverse-MSFE combination weights generate mean (median) monthly portfolio returns of 1.7% (1.6%), and a remarkable Sharpe ratio of 0.36. Even though the 90% (empirical) confidence band for realized portfolio returns remains rather wide – spanning the interval -2.3% to 5.2% – this seems a quite remarkable performance. For comparison, a statistic random-walk based strategy would yield a mean monthly return of 1.2% only, a Sharpe ratio of 0.20, and imply an even wider 90% confidence band of [-4.5%, 5.3%]. De Miguel et al.'s (2006) "$1/N$" equally weighted portfolio gives a mean portfolio return of 0.8% per month and considerably lower Sharpe ratio, 0.16. Among

[32] We thank an anonymous referee for pointing out the usefulness of these further benchmarks. Notice that over horizons longer than one month, even simpler benchmarks such as "$1/N$" and "50-50" would actually imply the need to rebalance to take into account that as assets give different returns, their portfolio weights change accordingly. Clearly, "$1/N$" and "50-50" differ because the latter only invests in risky assets.

[33] Clearly, the Sharpe ratio and the mean-variance statistic are strictly related. The difference is the latter is more tightly connected with the objective function that the vector of optimal weights is maximizing.

the strategies that do not exploit combinations, the four-state model comes up on top, with a mean portfolio return of 1.1% that however represents satisfactory compensation for risk (the Sharpe ratio is 0.21) since realized returns are much less volatile when regimes are taken into account (e.g., the 90% confidence band is [-3.9%, 4.6%]). Forecast combinations also systematically maximize the sample analog of the mean-variance objective, equation (16).[34]

Although a difference of 0.5% in realized portfolio returns between the random walk benchmark and forecast combination-based strategies may appear small (but the difference in realized Sharpe ratio is rather large), there is in fact strong statistical evidence that a forecast combination scheme may produce superior portfolio outcomes than any other models, including those that take into account the presence of regimes. For instance, assuming $\lambda = 1$ and treating portfolio returns as the relevant loss function, a DM-type test of the null of equal portfolio performance between the inverse-MSFE scheme and the MSIH(4,0)-VAR(1) produces a p-value of 0.03.[35] When the relevant loss function is identified – probably more correctly – with the mean-variance statistic, the p-value is instead 0.001.[36] In short, with a small margin of type I error (at worst 6.7%), both portfolio returns and returns adjusted for variability are significantly higher under the best forecast combination schemes than under the best fitting and predicting models that capture regimes.[37]

What is then the price tag we may attach to the possibility of supplementing careful modeling of regime switches in predictability relationships involving US excess asset returns with forecast combination schemes? Our results show that – depending on the risk aversion coefficient – such a value ranges from a 8.4% increase in mean annualized portfolio returns to an increase in the realized Sharpe ratio from 0.20 to 0.36. Both measures seems hardly negligible and may represent a powerful incentive for money managers to seriously entertain pooling the predictions yielded by a number of competing models as a serious alternative to simply selecting whichever of the models may have guaranteed a superior performance in the past.

7.3. *Sub-Sample Performance*

On an intuitive dimension, regime switching models may improve out-of-sample performance – both at forecasting asset returns and in realized portfolio terms – because they allow for different estimated models of the relationship between asset returns and predictors to characterize periods of bull and bear markets, see e.g. the discussion in Guidolin and Ono (2006). If this is the case, we should find that the performance of either regime switching models or models that pool forecasts produced by models that account for regime shifts should not be too sensitive to specific sub-periods of the overall (pseudo) out-of-sample interval, 1985-2004, provided that the periods are long enough to incorporate both bull and bear markets. Therefore we have divided the 20 years used in our experiment in four sub-samples, each including 5 years and consequently an adequate number of regime shifts (see Figure 3), and examined the performance of each of our nine models

[34]For comparison we have also computed out-of-sample performance statistics for the 100% stocks and bonds portfolios, respectively. The 100% stock portfolio gives a mean monthly return of 1.2% (i.e. the implied risk premium is 5.6% per year) and a Sharpe ratio of 0.20. The corresponding statistics for the 100% bond portfolio are 0.8% (the implied risk premium is 0.9% per year) and 0.11. In our sample, diversification obviously pays off on the basis of most econometric frameworks.

[35]The p-values are boostrapped using the scheme introduced in Section 5.2. The p-values for this null hypothesis obtained in the cases of $\lambda = 0.2$ and 2 are 0.02 and 0.07, respectively.

[36]The p-values for this null hypothesis obtained in the cases of $\lambda = 0.2$ and 2 are 0.00 and 0.003, respectively.

[37]It is instead more difficult to distinguish the two combination schemes – equally weighted and inverse-MSFE relative weights – in the sense that portfolio returns do not statistically exceed the ones from equal weights. However the null of equal mean-variance objectives may again be rejected for all levels of risk aversion.

along the usual dimensions. To save space, we report only results for the two most plausible risk aversion coefficients, $\lambda = 0.5$ and 1.

Table 10 reports the relevant results. Importantly, even though the specific performance measures prove to be highly time-varying (in particular, as one should expect, performances are excellent over 1995-1999 and rather disappointing for the periods 2000-2004), the relative ranking across models remains essentially identical to Table 9 for each of the four 5-year periods. The forecast combination-based portfolio strategies turn out to be rather special because they are the only ones to *always* generate positive Sharpe ratios, although the ratios vary from the 2000-2004 minimum of 0.07-0.09 (for inverse-MSFE portfolios) to 0.43-0.62 in the booming 1995-1999 years.[38]

Table 10 approximately here

7.4. *Transaction Costs*

One further dimension in which alternative investment strategies may differ is for their implication for the portfolio turnover, i.e. the relative frequency (and size) at which portfolio weights are changed over time, as a result of model-specific revisions of expected returns, variances, and covariances. Clearly, turnover may actually affect out-of-sample portfolio performance when there are non-negligible transaction costs which penalize the decision to change portfolio structures. In this respect, while a few of our benchmarks (such as the "$1/N$" and "50-50" portfolios) imply zero turnover and therefore no transaction costs, a few of the other models (especially, regime switching models and forecast combination-based portfolios which also involve regime switching forecasts) may in principle be quite expensive, as they require active portfolio rebalancing. We therefore proceed to recompute monthly means and medians of portfolio returns, Sharpe ratios, and realized mean-variance utility in the presence of transaction costs, tc. We specify the transaction cost function after Lynch and Balduzzi (2000):

$$tc_t = \phi_p \sum_{j=1}^{3} |\hat{w}_t^j - \hat{w}_{t-1}^j| + \phi_f I_{max_j |\hat{w}_t^j - \hat{w}_{t-1}^j| \neq 0},$$

where $\hat{w}_t^j$ is the optimal portfolio weight of asset j at time t, $I_{max_j |\hat{w}_t^j - \hat{w}_{t-1}^j| \neq 0}$ is an indicator function that takes unit value when *any* of the differences $\hat{w}_t^j - \hat{w}_{t-1}^j$ is non-zero (i.e., the weight of asset j has changed between month $t-1$ and t). ϕ_p and ϕ_f are two parameters which measure the unit incidence of transaction costs. In what follows we specify $\phi_p = 0.25\%$ and $\phi_f = 0.1\%$.[39] In essence, instead of computing and reporting W_{t+1}, we focus on $W_{t+1} - tc_{t+1}$.

Table 11 shows the results for the two most realistic values of λ, 0.5 and 1. The realized out-of-sample performance measures ought to be compared to the ones in Table 9. As one should expect, the statistics for

[38] The only minor difference (of no relevance to risk-averse investors) is that periods can be found for which either Neeley-Weller's or Pesaran-Timmermann's models yield the highest mean portfolio returns.

[39] These are the baseline values used in Lynch and Balduzzi (2000, pp. 2296-2297). The fixed cost parameter translates into paying a fee of \$10 whenever a \$100,000 portfolio is reshuffled. The proportional cost parameter implies a roundtrip cost of 0.5 percent. For instance, the Fidelity's Spartan Total Market Index Fund is an index fund that attempts to track the value of a U.S. value-weighted equity portfolio and it has been charging a redemption fee of 0.5 percent on fund shares sold within three and six months of purchase, respectively. Although it is possible to think that equity portfolios may implier higher values for ϕ_p and ϕ_f, the opposite applies to bond and especially cash portfolios (which are almost free). On average, this implies that the assumed values for ϕ_p and ϕ_f are quite realistic.

the benchmarks are not influenced by transaction costs. Interestingly, because these strategies imply that for long periods corner solutions (in which 100% of the portfolio is destined to either stocks or bonds) are obtained, the performance of Neely-Weller and Pesaran-Timmermann strategies are hardly affected. Finally, the performance of the regime switching models is substantially worsened when tc is taken into account. For instance, under $\lambda = 1$, the mean monthly return of the MSIH(4,0)-VAR(1) model goes from 1.12% to 0.88%; correspondingly the Sharpe ratio declines from 0.211 to 0.162. In spite of this, the out-of-sample performance of forecast-combination driven strategies remains overwhelmingly superior. Once more, assuming $\lambda = 1$, gives a Sharpe ratio of 0.28 for the inverse-MSFE strategy vs. 0.16 for the zero-transaction cost portfolios and a 0.20 for Neeley-Weller and Pesaran-Timmermann's strategies.

Table 11 approximately here

8. **Conclusion**

This paper has investigated two closely related issues. First, we have proposed to use multivariate regime switching models to capture the presence of time-variation – structural shifts – in the predictability patterns involving US asset returns and macroeconomic variables. Using a long monthly data set (1926-2004) we find overwhelming evidence of regimes in the joint process for returns and macroeconomic factors. The good performance of our four-state model at fitting the entire density of the data and its forecasting performance stress that payoffs may exist in explicitly modeling the presence of regimes. Second, we have built a rather strong case – both in statistical and economic terms – in favor of using forecast combinations in the presence of regime switching, i.e. against treating data sets in which regimes are detected with the classical "horse race" approach that aims at picking the best performing model(s) for forecasting purposes. On the opposite, our results indicate that pooling schemes exist that both improve the prediction performance in (pseudo) out-of-sample experiments, and that are of help to portfolio managers.

Interestingly, our results have also provided confirmation for a phenomenon often noted in the forecasting literature: especially in the case of non-stationary data, to try to run regression and estimate optimal combination weights may be foolish. The underlying process is likely to be so complicated that the OLS estimates are ridden of estimation error and as such hardly of any use. As our empirical analysis shows, the smart choice becomes then to either adopt pooling schemes that refrain from estimation or – even better – adopt mixed schemes that minimize the estimation needs and therefore maximize the chances of producing accurate forecasts, such as Stock and Watson's (2003) relative performance weights.

Several extensions of this paper appear to be rather natural. First of all, there is nothing compelling about using multivariate regime switching models to study time-varying linkages between financial markets and the macroeconomy. Other modeling approaches might prove useful, see e.g. Ravazzolo, Paap, van Dijk, and Franses (2006). Second, for simplicity we have computed combination weights and assessed prediction accuracy using an essentially univariate approach, in which – although some of the models under consideration are multivariate in nature – the focus has been mainly directed to forecasting excess stock and bond returns, one at the time. It would be interesting to consider whether adopting an explicit multivariate approach may change the results found here. Third, there is ongoing theoretical research on the properties of alternative relative performance schemes, which implies that our paper is likely to provide at best a lower bound to the gains that may be delivered by "smartly" selected pooling designs. For instance, Yang (2004) studies (Bayesian) recursive forecast combination algorithms based on relative performance measures. The advantage of these methods is ensure some degrees of minimal performance relative to the best (yet unknown) forecasting models. Finally, our application in Section 7 to portfolio management merely scratches the surface of the relevant issues. For instance, other natural loss functions may be involved by the practice of portfolio management for which recent research has revealed that modeling non-stationarities may be crucial (see e.g., Guidolin and Timmermann (2006b) for applications to value-at-risk computations).

References

[1] Aiolfi, M., and Timmermann, A. (2005). Persistence of Forecasting Performance and Combination Strategies. *Journal of Econometrics*, forthcoming.

[2] Ang A., and Bekaert, G. (2002a). International Asset Allocation with Regime Shifts. *Review of Financial Studies* 15, 1137-1187.

[3] Ang, A., and Bekaert, G. (2002b). Regime Switches in Interest Rates. *Journal of Business and Economic Statistics* 20, 163-182.

[4] Bates, J., and Granger, C. (1969). The Combination of Forecasts. *Operations Research Quarterly* 20, 451-468.

[5] Bekaert, G., and Hodrick, R. (1992) Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets. *Journal of Finance* 47, 467-509.

[6] Berkowitz, J. (2001). Testing Density Forecasts with Applications to Risk Management. *Journal of Business and Economic Statistics* 19, 465-474.

[7] Bossaerts, P., and Hillion, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Review of Financial Studies* 12, 405-428.

[8] Bunn, D. (1985). Statistical Efficiency in the Linear Combination of Forecasts. *International Journal of Forecasting* 1, 151-163.

[9] Campbell, J. (1987). Stock Returns and the Term Structure. *Journal of Financial Economics* 18, 373-399.

[10] Chan, L., Karceski, J., and Lakonishok, J. (1998). The Risk and Return from Factors. *Journal of Financial and Quantitative Analysis* 33, 159-188.

[11] Clemen, R. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting* 5, 559-583.

[12] Clements M., and Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge University Press.

[13] Clements, M., and Krolzig, H. (1998). A Comparison of the Forecast Performance of Markov-Switching and Threshold Autoregressive Models of US GNP. *The Econometrics Journal* 1, C47-C75.

[14] Clements, M., and Smith, J. (1997). The Performance of Alternative Forecasting Methods for SETAR Models. *International Journal of Forecasting* 13, 463-475.

[15] Cutler, D., Poterba, J., and Summers, L. (1989). What Moves Stock Prices? *Journal of Portfolio Management* 15, 4-12.

[16] Davies, R. (1977). Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative. *Biometrika* 64, 247-254.

[17] Diebold, F. (1988). Serial Correlation and the Combination of Forecasts. *Journal of Business and Economic Statistics* 6, 105-111.

[18] Diebold, F., Gunther, T., and Tay, A. (1998). Evaluating Density Forecasts. *International Economic Review* 39, 863-883.

[19] Diebold, F., and Lopez, J. (1996). Forecast Evaluation and Combination. In G., Maddala, and C., Rao, (Eds.), *Handbook of Statistics 14: Statistical Methods of Finance,* Elsevier, 241-268.

[20] Diebold, F. , and Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

[21] Diebold, F., Rudebusch, G., and Sichel, D. (1993). Further Evidence on Business Cycle Duration Dependence. In J., Stock and M., Watson (Eds.), *Business Cycles, Indicators, and Forecasting*, University of Chicago Press and NBER, 255-280.

[22] Dunis, C., Laws, J., and Chauvin, S. (2001). The Use of Market Data and Model Combinations to Improve Forecast Accuracy. In Dunis, C., A., Timmermann, and J., Moody (Eds.), *Developments in Forecast Combination and Portfolio Choice*, John Wiley and Sons, Inc., 45-80.

[23] Elliott, G., and Timmermann, A. (2004). Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions. *Journal of Econometrics* 122, 47-79.

[24] Fama, E., and French, K. (1988). Dividend Yields and Expected Stock Returns. *Journal of Financial Economics* 19, 3-29.

[25] Fama, E., and French, K. (1989). Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics* 25, 23-49.

[26] Fama, E., and French, K. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33, 3-36.

[27] Fama, E., and Schwert, W. (1977). Asset Returns and Inflation. *Journal of Financial Economics* 5, 115-146.

[28] Fenton, V., and Gallant, R. (1996). Qualitative and Asymptotic Performance of SNP Density Estimation. *Journal of Econometrics* 74, 77-118.

[29] Fugazza, C., Guidolin, M., and Nicodano, G. (2006). Investing for the Long-Run in European Real Estate. *Journal of Real Estate Finance and Economics,* forthcoming.

[30] Garcia, R. (1998). Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models. *International Economic Review* 39, 763-788.

[31] Goyal, A., and Welch, I. (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science* 49, 639-654.

[32] Gray, S. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics* 42, 27-62.

[33] Guidolin, M., and Ono, S. (2006). Are the Dynamic Linkages Between the Macroeconomy and Asset Prices Time-Varying? *Journal of Economics and Business* 58, 480-518.

[34] Guidolin, M., and Timmermann, A. (2005a). Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns. *Economic Journal* 115, 111-143.

[35] Guidolin, M., and Timmermann, A. (2005b). Optimal Forecast Combination Weights under Regime Shifts: An Application to US Interest Rates. Mimeo, Federal Reserve Bank of St. Louis and UCSD.

[36] Guidolin, M., and Timmermann, A. (2006a). An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns. *Journal of Applied Econometrics* 21, 1-22.

[37] Guidolin, M., and Timmermann, A. (2006b). Term Structure of Risk under Alternative Econometric Specifications. *Journal of Econometrics* 131, 285-308.

[38] Guidolin, M., and Timmermann, A. (2006c). Asset Allocation under Regime Switching. *Journal of Economic Dynamics and Control*, forthcoming.

[39] Hamilton, J. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57, 357-384.

[40] Hamilton, J., and Susmel, R. (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics* 64, 307-333.

[41] Hendry, D., and Clements, M. (2002). Pooling of Forecasts. *Econometrics Journal* 7, 1-31.

[42] James, C., Koreisha, S., and Partch, M. (1985). A VARMA Analysis of Casual Relations Among Stock Returns, Real Output, and Nominal Interest Rates. *Journal of Finance* 40, 1375-1384.

[43] Jorion, P. (1985). Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis* 21, 279-92.

[44] Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions*. Berlin, Springer-Verlag.

[45] Lynch, A., and Balduzzi, P. (2000). Predictability and Transaction Costs: The Impact on Rebalancing Rules and Behavior. *Journal of Finance* 55, 2285-2309.

[46] Neely, C., and Weller, P. (2000). Predictability in International Asset Returns: Reexamination. *Journal of Financial and Quantitative Analysis* 35, 601-620.

[47] Palm, F., and Zellner, A. (1992). To Combine or Not to Combine? Issues of Combining Forecasts. *Journal of Forecasting* 11, 687-701.

[48] Paye, B., and Timmermann, A. (2006). Instability of Return Prediction Models. *Journal of Empirical Finance*, forthcoming.

[49] Pesaran, H., and Timmermann, A. (1995). Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance* 50, 1201-1228.

[50] Rapach, D., and Wohar, M. (2006). In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining. *Journal of Empirical Finance* 13, 231-247.

[51] Ravazzolo, F., Paap, R., van Dijk, D., and Franses, P., H. (2006). Bayesian Model Averaging in the Presence of Structural Breaks. Mimeo, Erasmus University Rotterdam.

[52] Sola, M., and Driffill, J. (1994). Testing the Term Structure of Interest Rates Using a Stationary Vector Autoregression with Regime Switching. *Journal of Economic Dynamics and Control* 18, 601-628.

[53] Stock, J., and Watson, M. (2001). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In R., Engle, and H., White (Eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive Granger*, Oxford University Press, 1-44.

[54] Stock, J., and Watson, M. (1989). New Indexes of Coincident and Leading Economic Indicators. *NBER Macro-Economic Annual*, 351-409.

[55] Stock, J., and Watson, M. (2003). Forecasting Output and Inflation: the Role of Asset Prices. *Journal of Economic Literature* 41, 788-829.

[56] Timmermann, A. (2005). Forecast Combinations. Forthcoming in *Handbook of Economic Forecasting*. Elsevier Press.

[57] Turner, C., Starz, R., and Nelson, C. (1989). A Markov Model of Heteroskedasticity, Risk, and Learning in the Stock Market. *Journal of Financial Economics* 25, 3-22.

[58] De Miguel, V., Garlappi, L., and Uppal, R. (2006). 1/N. Mimeo, London Business School.

[59] Winkler, R. (1989). Combining Forecasts: A Philosophical Bias and Some Current Issues. *International Journal of Forecasting* 5, 605-609.

[60] Winkler, R., and Makridakis, S. (1983). The Combination of Forecasts. *Journal of the Royal Statistical Society Series A* 146, 150-157.

[61] Yang, Y. (2004). Combining Forecasts Procedures: Some Theoretical Results. *Econometric Theory* 20, 176-190.

Figure 1

Quantile-Quantile Plots (vs. a Gaussian with identical mean and variance) for Transformed z-Scores from Three-State VAR(1) Switching Model

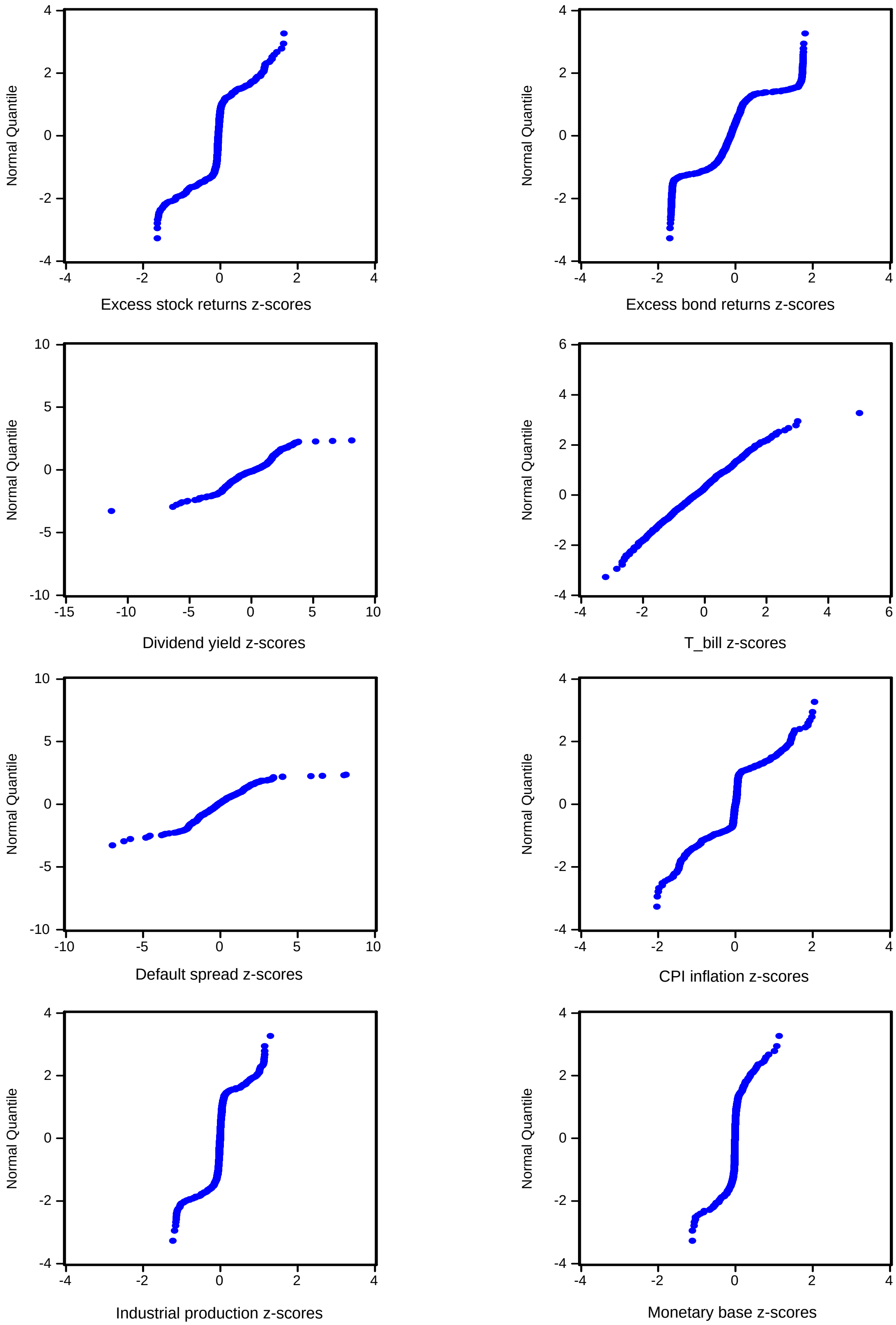


Figure 2

Quantile-Quantile Plots (vs. a Gaussian with identical mean and variance) for Transformed z-Scores from Four-State Model with Time-Invariant VAR(1) Matrix

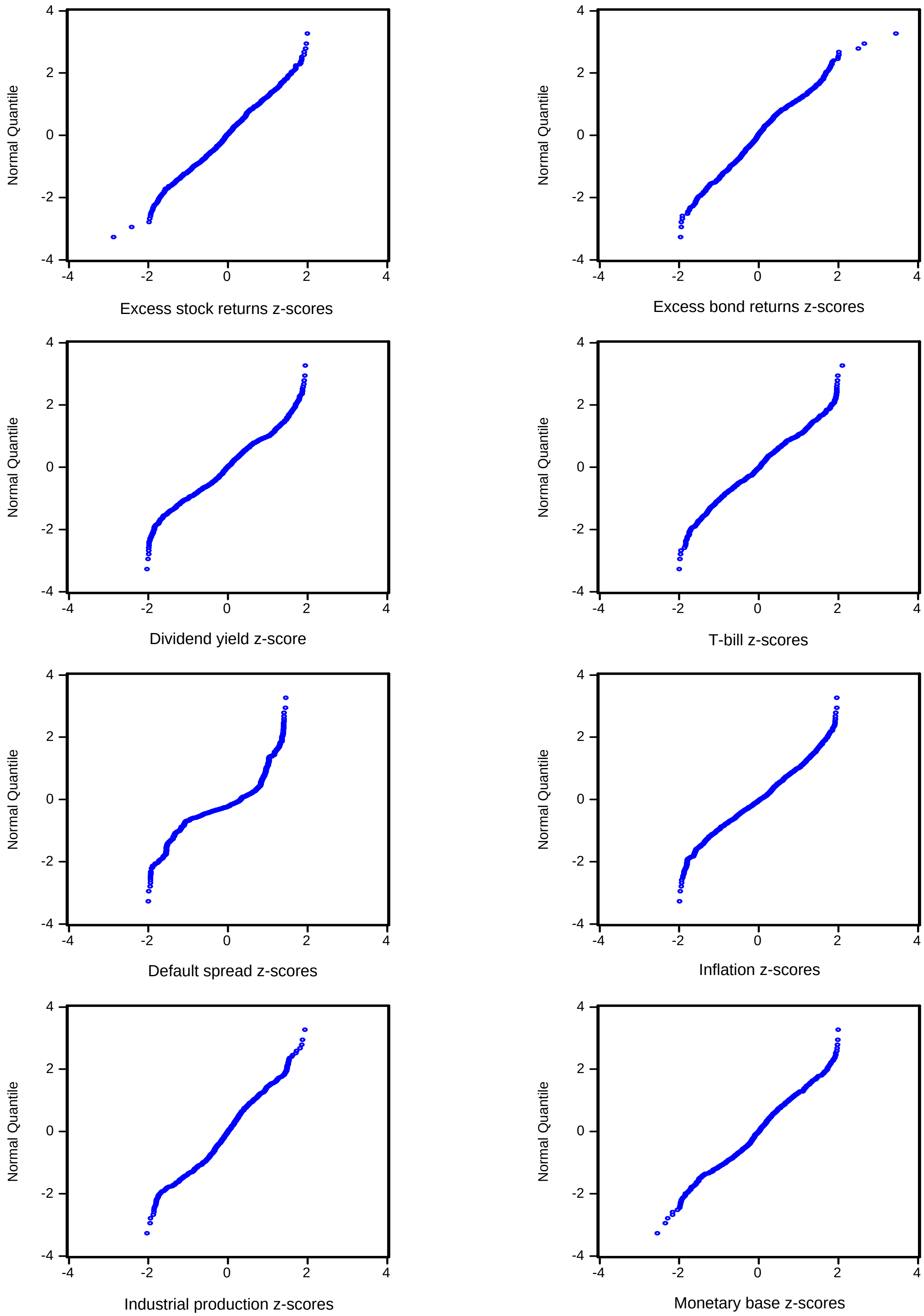


Figure 3

Smoothed Probabilities from Four-State Model

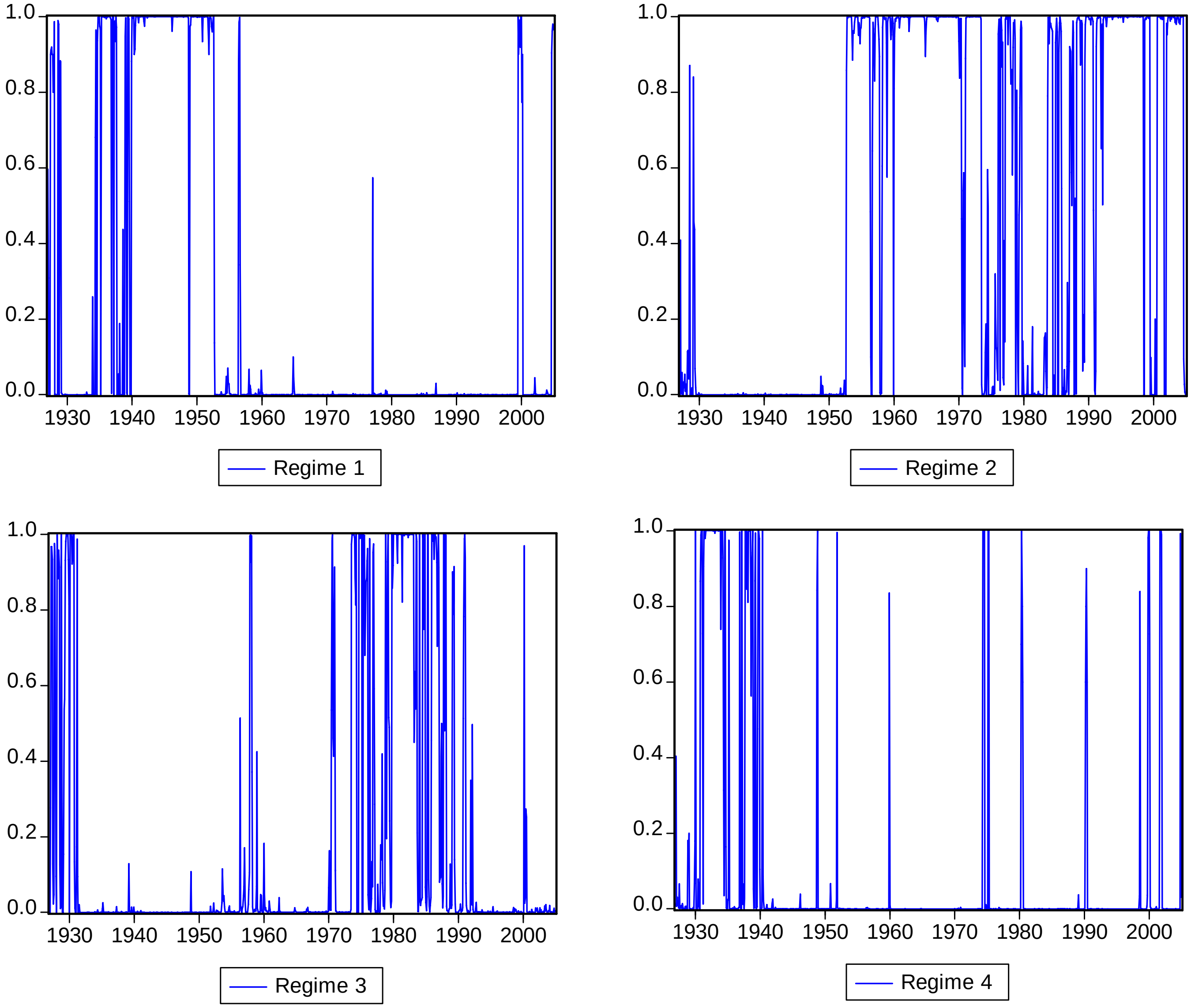


Figure 4

Comparing Fitted and Realized Values: Four-State Model with Time-Invariant VAR(1) Matrix

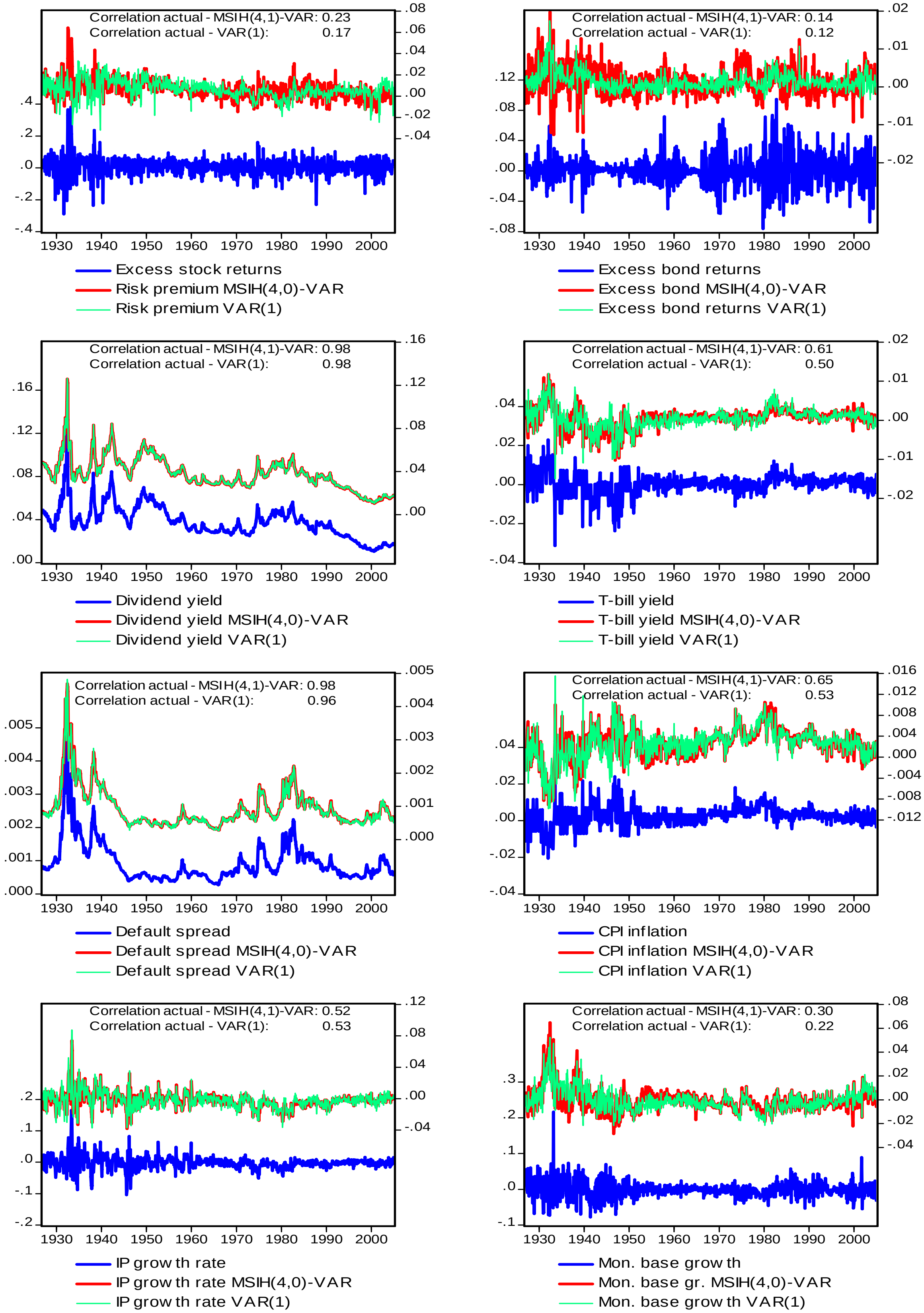
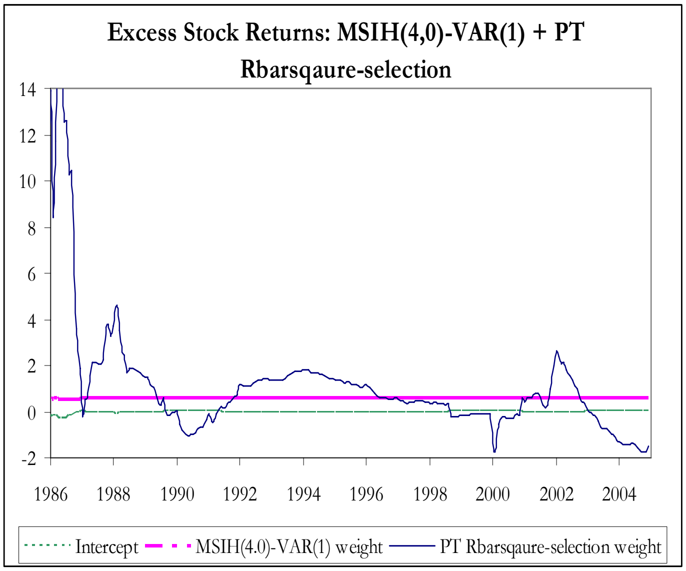
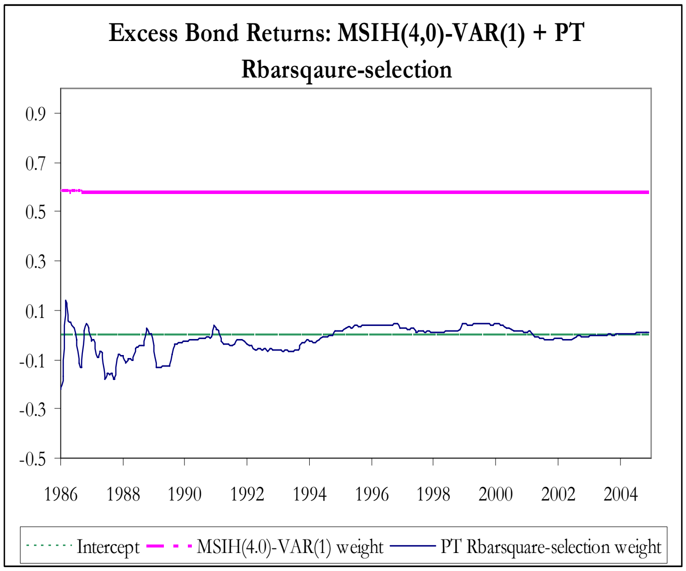


Figure 5

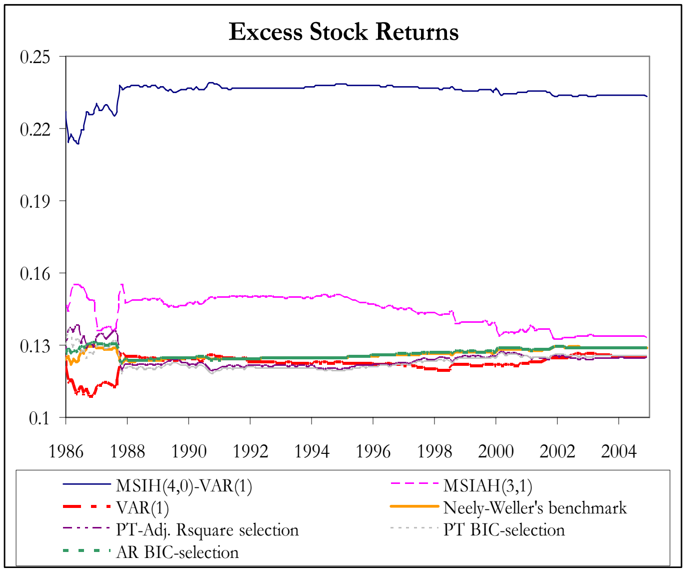
Recursive Combination Weights: One-Month Horizon

Panel A: OLS weights





Panel B: Inverse-MSFE weights



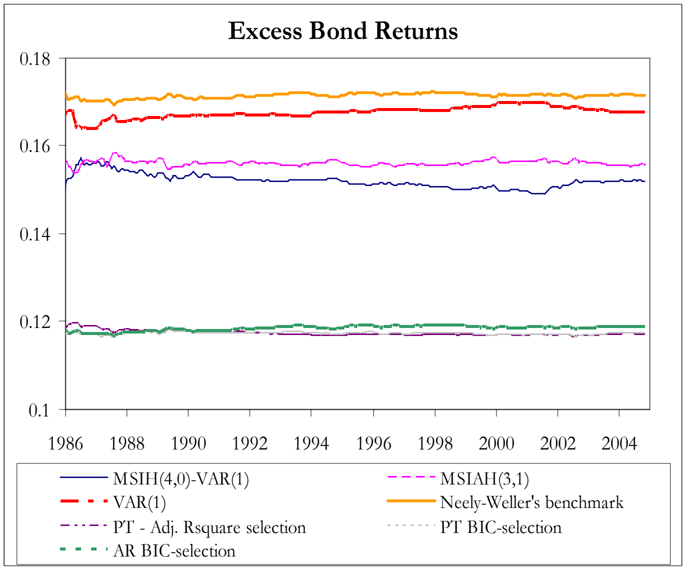
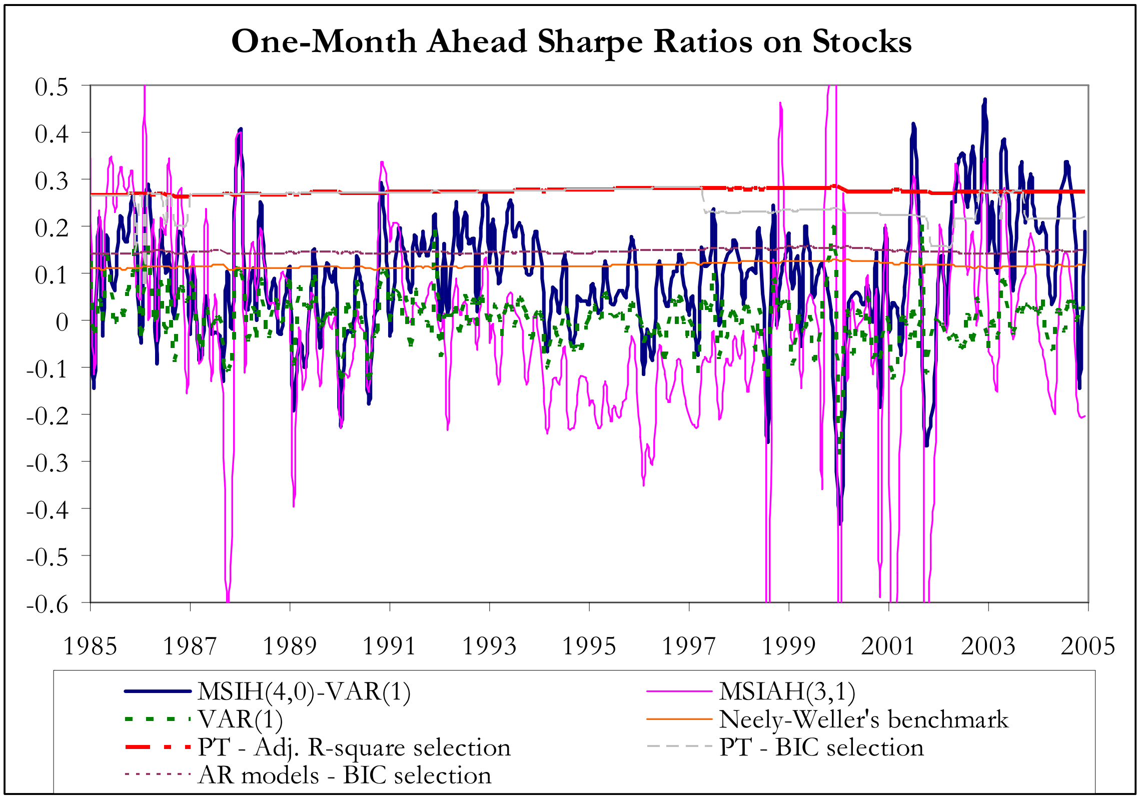


Figure 6

Implied Monthly Predicted Sharpe Ratios from Different Models



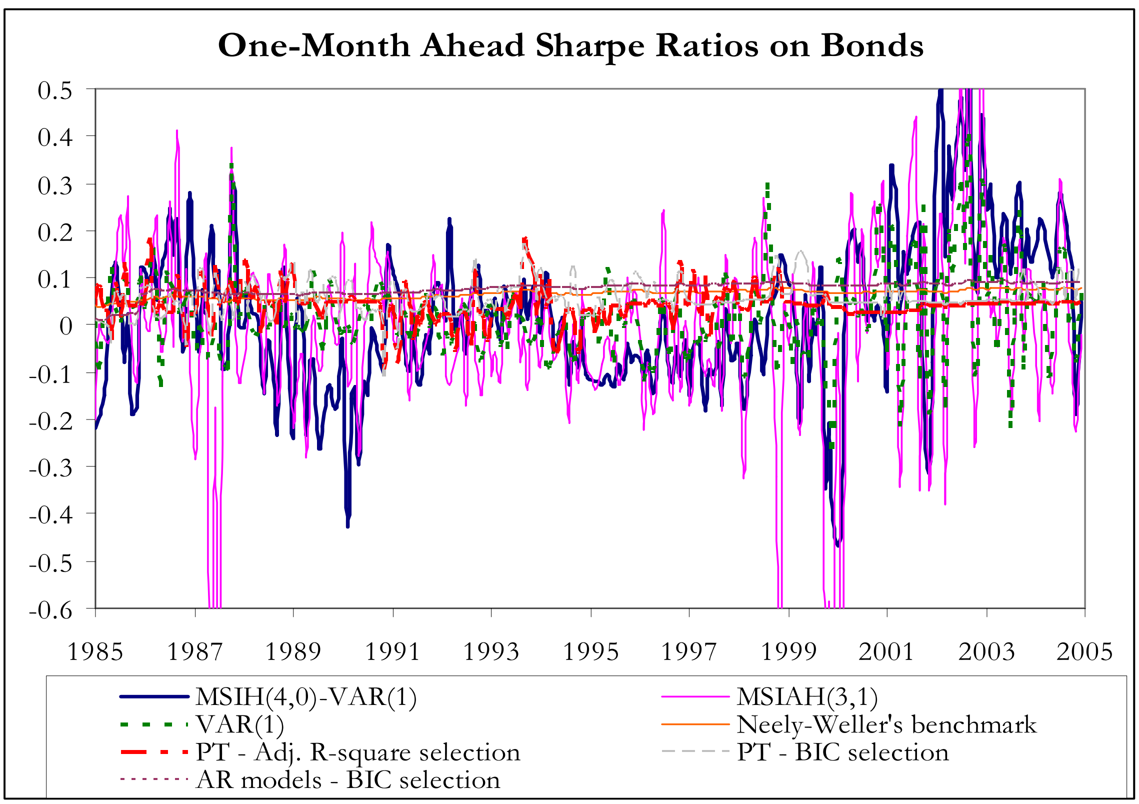


Table 1

Summary Statistics for Excess Stock and Bond Returns vs. Prediction Variables

The table reports a few summary statistics for monthly CRSP excess stock and (long-term) government bond return series, and macroeconomic variables employed as predictors of excess asset returns. Excess returns are calculated by difference with 30-day T-bill yields. The sample period is 1926:12 – 2004:12. In the case of equities, the CRSP universe spans stocks listed on the NYSE, the NASDAQ, and the AMEX. Data on bond returns refer to the CRSP 10-Year Treasury benchmark. All returns are expressed in monthly percentage terms. LB(j) denotes the j-th order Ljung-Box statistic.

Series	Mean	Median	St. Dev.	Skewness	Kurtosis	Jarque-Bera	LB(4)	LB(4)-squares
	Excess Asset Returns (Risk Premia)							
Value-weighted excess stock returns	0.6482	0.9900	5.4946	0.2133	10.6124	2269**	21.716**	166.87**
Excess bond returns (term premium)	0.1447	0.1400	1.8808	0.2447	5.5932	271.9**	5.1774	176.31**
	Prediction Variables							
12-month cumulated dividend yield	3.8132	3.6340	1.4987	0.9542	5.8183	452.3**	3334**	2829**
Real 1-month T-bill yield	0.0540	0.0700	0.5114	-1.9764	21.0381	13313**	542.13**	79.833**
Default spread	0.0943	0.0730	0.0600	2.4203	11.3805	3657**	3284**	2683**
CPI inflation rate	0.2498	0.2659	0.5279	1.1840	16.7930	7647**	596.9**	82.741**
Industrial production growth rate	0.2101	0.2270	2.0208	0.7663	13.2813	4219**	268.7**	372.7**
Real adj. monetary base growth rate	0.0381	0.1540	2.3031	1.7034	30.7269	30468**	34.722**	79.514**

* denotes 5% significance, ** significance at 1%.

Table 2 – part a

Estimates of a Four-State Switching Model with Time-Invariant VAR(1) Matrix

Panel A – Single State VAR(1) Model								
	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
1. Intercept	-0.3878	-0.0376	0.1022***	-0.0425	0.0001	0.0605	0.1356	0.5340*
	(0.6076)	(0.2092)	(0.0331)	(0.0522)	(0.0013)	(0.0518)	(0.1910)	(0.3037)
2. VAR(1) Matrix								
Stock excess returns	0.1046***	0.1108	0.2653*	-0.2514	0.5962	-0.3426	0.0653	0.1257**
	(0.0333)	(0.0960)	(0.1430)	(0.7425)	(3.5335)	(0.7735)	(0.0974)	(0.0586)
Bond excess returns	-0.0260**	0.0626*	-0.0020	0.0522	2.2183*	-0.0485	0.0115	-0.0104
	(0.0115)	(0.0331)	(0.0492)	(0.2557)	(1.2168)	(0.2664)	(0.0335)	(0.0202)
Dividend yield	-0.0060***	-0.0011	0.9835***	-0.0403	-0.3175*	-0.0285	-0.0029	-0.0052*
	(0.0018)	(0.0052)	(0.0078)	(0.0405)	(0.1925)	(0.0421)	(0.0053)	(0.0032)
T-bill real yield	-0.0011	0.0013	-0.0249*	0.5768***	1.2793***	0.1068*	-0.0518***	-0.0027
	(0.0029)	(0.0083)	(0.0123)	(0.0638)	(0.3037)	(0.0605)	(0.0084)	(0.0050)
Default spread	-0.0009***	0.0007***	0.0007**	0.0021	0.9628***	0.0002	-0.0005**	0.0000
	(0.0001)	(0.0002)	(0.0003)	(0.0016)	(0.0077)	(0.0017)	(0.0002)	(0.0001)
Inflation	0.0013	-0.0087	0.0232*	0.3867***	-1.2891***	0.8632***	0.0526***	0.0026
	(0.0028)	(0.0082)	(0.0122)	(0.0633)	(0.3011)	(0.0659)	(0.0083)	(0.0050)
IP real growth	0.0760***	-0.0372	-0.0854*	-0.9677***	3.3244***	-0.9506***	0.4008***	0.0396**
	(0.0105)	(0.0302)	(0.0449)	(0.2334)	(1.1108)	(0.2432)	(0.0306)	(0.0184)
Money real growth	-0.0025	0.0219	-0.2341***	-1.3385***	10.338***	-2.3610***	-0.0316	-0.1880***
	(0.0188)	(0.0543)	(0.0809)	(0.4201)	(1.9992)	(0.4377)	(0.0551)	(0.0332)
3. Correlations/Volatilities								
Stock excess returns	0.0542***							
Bond excess returns	0.1377**	0.0187***						
Dividend yield	-0.8830***	-0.1368**	0.0030**					
T-bill real yield	-0.0333	0.0561	0.0404	0.0047***				
Default spread	-0.2596***	0.0712	0.3324***	0.0371	0.0001*			
Inflation	0.0221	-0.0582	-0.0313	-0.9909***	-0.0298	0.0046***		
IP real growth	0.1128**	-0.0104	-0.1012**	0.4401***	-0.1744**	-0.4434***	0.0170***	
Money real growth	-0.0034	0.0594*	-0.0027	0.2329**	0.0677*	-0.2332**	0.0259	0.0031**
Panel B – Four State Model								
	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
1. Intercept								
Bull-rebound	0.8945**	0.2530***	0.0382*	-0.2271***	0.0022	0.2321**	0.1633*	-0.9862**
	(0.4420)	(0.1031)	(0.0185)	(0.0781)	(0.0046)	(0.1181)	(0.1090)	(0.3959)
Stable-growth	0.6454**	0.2955***	0.0169	-0.0008	0.0009	0.0225	0.2330*	-0.0108
	(0.2707)	(0.0957)	(0.0173)	(0.0416)	(0.0019)	(0.0418)	(0.1336)	(0.2049)
Expansion-peak	0.5782***	1.0141	0.0408*	0.0015	0.0032	0.0399	-0.0207	0.0478*
	(0.0874)	(0.5851)	(0.0206)	(0.0776)	(0.0019)	(0.0777)	(0.2502)	(0.0261)
Bear-recession	-1.9266***	-0.2598***	0.1301***	0.1575***	0.0089*	-0.1582**	0.4802***	0.4399***
	(0.4820)	(0.0570)	(0.0095)	(0.0191)	(0.0052)	(0.0793)	(0.1830)	(0.0836)
2. VAR(1) Matrix								
Stock excess returns	-0.0009	0.1057	0.0870	-1.3822***	4.7186***	-2.1260***	-0.0433**	0.0821*
	(0.0287)	(0.0719)	(0.1220)	(0.4207)	(0.0805)	(0.7250)	(0.0207)	(0.0450)
Bond excess returns	-0.0341*	0.0167	-0.0401*	-0.9605***	2.9146***	-1.0749***	-0.0079	0.0159
	(0.0173)	(0.0324)	(0.0242)	(0.2579)	(1.0360)	(0.3645)	(0.0204)	(0.0137)
Dividend yield	-0.0003	-0.0034	0.9940***	0.0343	-0.3475***	0.0619*	0.0032	-0.0044*
	(0.0011)	(0.0025)	(0.0044)	(0.0298)	(0.0753)	(0.0302)	(0.0038)	(0.0024)
T-bill real yield	0.0004	0.0045*	0.0011	0.4529***	0.3641**	0.0811	-0.0187	-0.0025
	(0.0025)	(0.0024)	(0.0102)	(0.0651)	(0.1875)	(0.0669)	(0.0092)	(0.0055)
Default spread	-0.0003*	0.0003*	0.0001	0.0034*	0.9640***	0.0031*	0.0000	-0.0000
	(0.0002)	(0.0002)	(0.0004)	(0.0022)	(0.0043)	(0.0020)	(0.0001)	(0.0001)
Inflation	-0.0006	-0.0092	-0.0008	0.4872***	-0.3839**	0.8599***	0.0190	0.0028*
	(0.0024)	(0.0054)	(0.0102)	(0.0655)	(0.1674)	(0.0672)	(0.0142)	(0.0016)
IP real growth	0.0235*	-0.0089	-0.0385	-0.8952	0.9575**	-0.9208***	0.3861***	0.0185
	(0.0120)	(0.0183)	(0.0362)	(0.2186)	(0.4569)	(0.2189)	(0.0326)	(0.0199)
Money real growth	0.0143	0.0233	0.0394	-1.4162***	6.8721***	-2.3985***	-0.0294	-0.1786***
	(0.0118)	(0.0253)	(0.0491)	(0.2968)	(1.9272)	(0.3015)	(0.0464)	(0.0324)

* denotes 10% significance, ** significance at 5%, *** significance at 1%.

Table 2 – part b

Estimates of a Four-State Switching Model with Time-Invariant VAR(1) Matrix

	Panel B – Four State Model							
	Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
3. Correlations/Volatilities								
Regime 1 (Bull-rebound):								
Stock excess returns	0.0385***							
Bond excess returns	0.1494**	0.0068***						
Dividend yield	-0.8806***	-0.1687***	0.0021***					
T-bill real yield	0.0438	0.0522	-0.0286	0.0071***				
Default spread	-0.1434**	-0.0127	0.1049**	-0.0593	3.9e-05*			
Inflation	-0.0429	-0.0522	0.0270	-0.9990***	0.0589	0.0071***		
IP real growth	-0.0918**	0.0388	0.0678*	0.5054***	-0.1040*	-0.5051***	0.0256***	
Money real growth	0.1742**	0.0220	-0.1864**	0.3201***	-0.0891*	-0.3211***	0.1508**	0.0351***
Regime 2 (Stable-growth):								
Stock excess returns	0.0365***							
Bond excess returns	0.1433**	0.0180***						
Dividend yield	-0.8999***	-0.1807***	0.0011***					
T-bill real yield	0.0784	0.0593	-0.0721	0.0023***				
Default spread	-0.0271	-0.0144	0.0204	0.0989*	3.98e-05*			
Inflation	-0.0917*	-0.0554	0.0884*	-0.9857***	-0.1034**	0.0023***		
IP real growth	0.0953*	-0.0492	-0.1158**	0.5290***	-0.0144	-0.5406***	0.0083***	
Money real growth	0.0539	0.0009	-0.0449	0.3928***	0.0472	-0.3953***	0.2029**	0.0108***
Regime 3 (Expansion-peak):								
Stock excess returns	0.0606***							
Bond excess returns	0.1965***	0.0277***						
Dividend yield	-0.9691***	-0.2080***	0.0026***					
T-bill real yield	0.0063	0.0896	-0.0427	0.0043***				
Default spread	0.0773*	0.3746***	-0.0809**	0.0494	0.0001**			
Inflation	-0.0665	-0.1141*	0.1161**	-0.9535***	-0.0045	0.0041***		
IP real growth	0.1126**	-0.0253	-0.1372**	0.5943***	-0.0540	-0.5789***	0.0125***	
Money real growth	-0.0977**	0.0234	0.0702*	0.3791***	-0.0690	-0.4093***	0.2576**	0.0198***
Regime 4 (Bear-recession):								
Stock excess returns	0.1196***							
Bond excess returns	0.1040*	0.0197***						
Dividend yield	-0.9225***	-0.1873***	0.0082***					
T-bill real yield	-0.1698**	0.1577**	0.1597**	0.0067***				
Default spread	-0.4836***	-0.1435**	0.4864***	0.0744	0.0003***			
Inflation	0.1727**	-0.1605**	-0.1631**	-0.9981***	-0.0797*	0.0067***		
IP real growth	0.3401***	0.0683	-0.2432***	0.1544**	-0.4347***	-0.1601**	0.0317***	
Money real growth	-0.0456	0.2287***	0.0456	0.0262	0.1042*	-0.0180	-0.1768**	0.0769***

4. Transition probabilities	Bull-rebound	Stable-peak	Expansion	Bear-recession
Bull-rebound	0.8939***	0.0237*	0.0207	0.0616
	(0.1560)	(0.0128)	(0.0249)	
Stable-growth	0.0100	0.9311***	0.0523**	0.0066
	(0.0349)	(0.2305)	(0.2410)	
Expansion-peak	0.0230	0.1517***	0.8030***	0.0381
	(0.2340)	(0.0555)	(0.1159)	
Bear-recession	0.1683***	0.0340*	0.0381**	0.7596
	(0.0200)	(0.0183)	(0.0160)	

* denotes 10% significance, ** significance at 5%, *** significance at 1%.

Table 3

Out-of-Sample, Recursive Predictive Performance

The table reports the root-mean-square forecast error, the predictive bias, and the forecast error standard deviation for the seven models described in the main text. The (pseudo) out-of sample period is 1985:01 – 2004:12. The models are recursively estimated on expanding windows 1926:12 – 1985:01, 1926:12 – 1985:02, up to 1926:12 – 2004:12 - *h*.

		Stock	Bond	Div. yield	T-bill	Default	Inflation	Growth	Money
One-month horizon		**Four-state MSIH(4,0)-VAR(1)**							
	Root-MSFE	3.773	1.285	1.443	0.880	0.152	0.277	1.385	2.421
	Bias	0.369	0.218	-1.428	-0.128	0.113	-0.224	0.359	0.561
	St. dev.	3.755	1.266	0.208	0.871	0.102	0.163	1.338	2.355
		Three-state MSIAH(4,1)							
	Root-MSFE	4.310	1.287	3.120	0.888	0.115	0.927	1.422	5.570
	Bias	0.883	0.287	2.533	-0.145	0.071	-0.167	0.318	0.266
	St. dev.	4.219	1.255	1.821	0.876	0.090	0.912	1.386	5.564
		Single-state VAR(1)							
	Root-MSFE	5.256	1.257	1.580	0.900	0.115	0.253	2.442	2.964
	Bias	0.631	0.227	-1.333	-0.124	0.071	-0.080	0.408	0.282
	St. dev.	5.218	1.236	0.848	0.891	0.090	0.240	2.408	2.951
		Neely and Weller's benchmark							
	Root-MSFE	6.138	1.238	3.163	0.891	0.115	0.923	3.438	5.881
	Bias	0.036	0.146	2.578	-0.144	0.071	-0.169	0.456	0.599
	St. dev.	6.138	1.229	1.832	0.879	0.091	0.907	3.408	5.850
		Pesaran and Timmermann – Adj. R^2 selection							
	Root-MSFE	4.485	2.152	2.114	0.274	0.006	0.266	0.893	4.311
	Bias	0.031	0.228	-0.008	0.000	-0.000	-0.002	0.009	-0.016
	St. dev.	4.485	2.140	2.114	0.274	0.006	0.266	0.893	4.311
		Pesaran and Timmermann – BIC selection							
	Root-MSFE	4.561	2.168	2.598	0.260	0.023	0.250	0.709	3.007
	Bias	-0.859	0.270	-2.451	0.0338	-0.006	0.085	-0.104	0.551
	St. dev.	4.480	2.151	0.861	0.258	0.022	0.235	0.702	2.956
		AR(p) univariate models (BIC Selection)							
	Root-MSFE	4.485	2.152	2.192	0.309	0.022	0.377	0.793	3.011
	Bias	-0.130	0.209	-2.017	-0.174	0.004	0.295	-0.370	-0.482
	St. dev.	4.483	2.142	0.858	0.256	0.022	0.235	0.702	2.972
One-year horizon		**Four-state MSIH(4,0)-VAR(1)**							
	Root-MSFE	4.560	2.129	0.478	0.272	0.017	0.252	0.739	3.001
	Bias	0.429	0.266	-0.297	0.058	-0.006	-0.063	-0.109	0.003
	St. dev.	4.540	2.112	0.374	0.266	0.016	0.243	0.731	3.001
		Three-state MSIAH(4,1)							
	Root-MSFE	5.076	2.385	0.830	0.300	0.022	0.291	0.935	3.347
	Bias	0.273	0.437	-0.628	-0.092	-0.010	0.081	-0.375	-0.318
	St. dev.	5.068	2.345	0.543	0.286	0.020	0.279	0.856	3.332
		Single-state VAR(1)							
	Root-MSFE	5.543	2.882	0.882	1.033	0.027	1.109	4.272	4.771
	Bias	-1.076	1.308	0.478	-0.816	0.001	0.894	-3.491	-2.762
	St. dev.	5.438	2.568	0.741	0.633	0.027	0.656	2.463	3.890
		Random Walk benchmarks							
	Root-MSFE	5.269	2.452	0.400	0.329	0.021	0.320	0.973	4.788
	Bias	0.007	0.181	-0.108	-0.025	-0.003	-0.002	0.017	-0.069
	St. dev.	5.269	2.445	0.385	0.328	0.021	0.320	0.973	4.788
		Pesaran and Timmermann – Adj. R^2 selection							
	Root-MSFE	4.695	2.184	2.395	0.278	0.023	0.264	0.763	3.102
	Bias	-0.411	0.214	-2.153	0.068	-0.010	-0.024	-0.227	-0.377
	St. dev.	4.677	2.173	1.047	0.269	0.021	0.263	0.728	3.078
		Pesaran and Timmermann – BIC selection							
	Root-MSFE	4.639	2.149	2.406	0.272	0.023	0.262	0.735	3.070
	Bias	-0.352	0.192	-2.192	0.062	-0.010	-0.018	-0.157	-0.442
	St. dev.	4.627	2.140	0.992	0.265	0.021	0.262	0.719	3.038
		AR(p) univariate models (BIC Selection)							
	Root-MSFE	4.650	2.164	2.163	0.267	0.021	0.258	0.728	3.138
	Bias	0.013	0.182	-1.961	0.034	-0.004	0.080	-0.040	-0.437
	St. dev.	4.650	2.156	0.912	0.264	0.021	0.245	0.727	3.107

Table 4

Predictive Accuracy Tests – 1-Month Horizon

The table reports Diebold-Mariano test statistics for pairwise comparisons of the MSE produced by different forecasting models. The test is applied to forecast errors from recursive, 1-step ahead forecasts of excess stock and bond returns using a variety of models. Numbers in bold highlight pairs of models for which one forecast function significantly outperforms the other at a 5% level using Diebold and Mariano (1995) asymptotic normal distribution. Numbers in parenthesis are p-values obtained from an application of the block bootstrap to MSE differentials (based on 50,000 independent block trials). Statistics illustrate the comparative forecasting performance of the model in the row vs. the model in the column: a negative (positive) value indicates that the row model out- (under-) performs the column model. Below the main diagonal, we present results for excess stock returns; above the main diagonal, results are for excess bond returns.

		PT – Adj. R^2 selection	PT – BIC selection	AR - BIC selection	Random walk benchmark	VAR(1)	MSIAH(3,1)	MSIH(4,0)-VAR(1)
		`Excess Bond Returns						
PT – Adj. R^2 selec`tion	Excess Stock Returns		**-2.719** (0.143)	0.126 (0.806)	**3.505** (0.006)	1.028 (0.278)	1.063 (0.327)	**7.273** (0.003)
PT – BIC selection		0.586 (0.471)		**2.175** (0.074)	**2.719** (0.042)	1.288 (0.251)	1.172 (0.105)	**7.521** (0.002)
AR - BIC selection		0.022 (0.098)	-1.579 (0.468)		**2.126** (0.181)	1.013 (0.346)	1.030 (0.043)	**7.222** (0.007)
Random walk		**3.405** (0.034)	1.299 (0.270)	0.022 (0.819)		-0.028 (0.845)	**-1.967** (0.144)	-0.294 (0.674)
VAR(1)		0.786 (0.410)	0.092 (0.167)	0.712 (0.404)	0.922 (0.317)		**-6.442** (0.038)	-0.956 (0.510)
MSIAH(3,1)		**-2.297** (0.048)	**-2.697** (0.123)	**-2.235** (0.080)	-0.998 (0.275)	**-2.002** (0.091)		0.125 (0.923)
MSIH(4,0)-VAR(1)		**-6.326** (0.000)	**-5.390** (0.021)	**-5.519** (0.046)	-1.817 (0.157)	**-6.275** (0.031)	**-2.255** (0.065)	

Table 5

Predictive Accuracy Tests – 12-Month Horizon

The table reports Diebold-Mariano test statistics for pairwise comparisons of the MSE produced by different forecasting models. The test is applied to forecast errors from recursive, 12-step ahead forecasts of excess stock and bond returns using a variety of models. Numbers in bold highlight pairs of models for which one forecast function significantly outperforms the other at a 5% level using Diebold and Mariano (1995) asymptotic normal distribution. Numbers in parenthesis are p-values obtained from an application of the block bootstrap to MSE differentials (based on 50,000 independent block trials). Statistics illustrate the comparative forecasting performance of the model in the row vs. the model in the column: a negative (positive) value indicates that the row model out- (under-) performs the column model. Below the main diagonal, we present results for excess stock returns; above the main diagonal, results are for excess bond returns.

		PT – Adj. R^2 selection	PT – BIC selection	AR - BIC selection	Random walk benchmark	VAR(1)	MSIAH(3,1)	MSIH(4,0)-VAR(1)
		`Excess Bond Returns						
PT – Adj. R^2 selec`tion	**Excess Stock Returns**		**2.640**	1.657	1.671	**-3.458**	-1.323	1.278
			(0.049)	(0.103)	(0.099)	(0.009)	(0.283)	(0.203)
PT – BIC selection		-1.237		0.330	1.352	**-3.565**	**-2.007**	1.818
		(0.274)		(0.794)	(0.169)	(0.034)	(0.060)	(0.084)
AR - BIC selection		-1.709	1.472		1.603	**-3.485**	-1.822	1.628
		(0.104)	(0.241)		(0.103)	(0.010)	(0.084)	(0.150)
Random walk		1.602	1.450	**2.968**		-1.825	**-4.854**	**3.850**
		(0.088)	(0.184)	(0.030)		(0.085)	(0.005)	(0.009)
VAR(1)		1.770	1.690	1.147	1.157		**4.367**	**3.647**
		(0.083)	(0.140)	(0.359)	(0.320)		(0.017)	(0.024)
MSIAH(3,1)		1.666	1.610	1.553	1.509	0.357		**3.586**
		(0.134)	(0.114)	(0.145)	(0.104)	(0.749)		(0.011)
MSIH(4,0)-VAR(1)		**-2.056**	**-1.979**	**-2.072**	-0.836	**-2.353**	-0.637	
		(0.059)	(0.063)	(0.047)	(0.291)	(0.050)	(0.536)	

Table 6

Recursive Predictive Performance of Forecast Combinations – One Month Horizon

The table reports the root-mean-square forecast error, the predictive bias, and the forecast error standard deviation for a number of alternative forecast combination schemes. The first panel concerns combinations in pairs, when a baseline forecast (i.e. forecast 1) is held fixed at the forecast function listed in the header. Pairwise weights are estimated by ordinary least squares. The second panel groups a few benchmarks: (i) when all forecast functions are recursively combined (with least-squares weights); (ii) when weights are fixed to be identical across all forecast functions; Stock and Watson's (2003) scheme in which weights are inversely proportional to the recursive sample MSFE of each forecast function. The (pseudo) out-of sample period is 1985:01 – 2004:11.

	Excess Stock Returns			**Excess Bond Returns**		
	Root-MSFE	**Bias**	**Std. Dev.**	**Root-MSFE**	**Bias**	**Std. Dev.**
Combination with:	**Panel A1: Baseline forecast: MSIH(4,0)-VAR(1)**					
MSIAH(3,1)	6.141	-0.006	6.141	2.812	0.005	2.812
VAR(1)	6.133	0.028	6.133	2.808	0.010	2.808
Random walk bench.	6.130	-0.015	6.130	2.810	-0.014	2.810
PT – Adj. R^2 selection	6.136	-0.013	6.136	2.807	0.010	2.806
PT – BIC selection	6.134	-0.007	6.134	2.807	0.009	2.807
AR - BIC selection	6.134	-0.031	6.134	2.811	-0.002	2.811
	Panel A2: Other benchmarks					
Baseline MSIH(4,0)-VAR(1)	3.773	0.369	3.755	1.285	0.218	1.266
All forecast functions	5.026	0.345	5.014	0.439	-0.014	0.439
Equally weighted	2.169	-0.303	2.148	1.047	0.106	1.042
Inv. MSFE-weighted	1.567	-0.301	1.538	0.689	0.057	0.687
	Panel B1: Baseline forecast: Pesaran and Timmermann – Adj. R^2 selection					
MSIAH(3,1)	6.096	-0.121	6.094	2.822	-0.004	2.822
VAR(1)	4.653	0.375	4.638	2.212	-0.165	2.206
Neely-Weller bench.	4.593	0.552	4.560	2.172	0.194	2.164
MSIH(4,0)-VAR(1)	6.136	-0.013	6.136	2.806	0.010	2.806
PT – BIC selection	4.646	0.168	4.643	2.129	-0.100	2.126
AR - BIC selection	4.588	0.153	4.586	2.173	0.089	2.171
	Panel B2: Other benchmarks					
Baseline PT – Adj. R^2 selection	4.485	0.031	4.627	2.152	0.228	2.140
All forecast functions	5.026	0.345	5.014	0.439	-0.014	0.439
Equally weighted	2.169	-0.303	2.148	1.047	0.106	1.042
Inv. MSFE-weighted	1.567	-0.301	1.538	0.689	0.057	0.687

Table 7

Recursive Predictive Performance of Forecast Combinations – 12-Month Horizon

The table reports the root-mean-square forecast error, the predictive bias, and the forecast error standard deviation for a number of alternative forecast combination schemes. The first panel concerns combinations in pairs, when a baseline forecast (i.e. forecast 1) is held fixed at the forecast function listed in the header. Pairwise weights are estimated by ordinary least squares. The second panel groups a few benchmarks: when all forecast functions are recursively combined (with least-squares weights); when weights are fixed to be identical across all forecast functions; Stock and Watson's (2003) scheme in which weights are inversely proportional to the recursive sample MSFE of each forecast function. The (pseudo) out-of sample period is 1985:01 – 2003:12.

	Excess Stock Returns			**Excess Bond Returns**		
Combination with:	**Root-MSFE**	**Bias**	**Std. Dev.**	**Root-MSFE**	**Bias**	**Std. Dev.**
	Panel A: Baseline forecast: MSIH(4,0)-VAR(1)					
MSIAH(3,1)	4.665	-0.041	4.665	2.262	0.031	2.262
VAR(1)	4.663	-0.082	4.662	2.240	-0.014	2.240
Neely-Weller bench.	4.561	-0.044	4.561	2.276	0.386	2.243
PT – Adj. R^2 selection	4.550	0.146	4.548	2.269	0.045	2.268
PT – BIC selection	4.653	0.033	4.653	2.262	0.015	2.262
AR - BIC selection	4.579	-0.037	4.579	2.265	0.357	2.237
	Panel A: Other benchmarks					
Baseline MSIH(4,0)-VAR(1)	4.560	2.129	4.540	2.129	0.266	2.112
All forecast functions	5.234	-1.340	5.060	3.277	-0.018	3.277
Equally weighted	4.485	-0.124	4.483	2.093	0.400	2.054
Inv. MSFE-weighted	4.100	-0.146	4.097	1.924	0.358	1.890
	Panel B: Baseline forecast: Pesaran and Timmermann – BIC selection					
MSIAH(3,1)	4.690	-0.087	4.689	2.232	0.031	2.232
VAR(1)	4.763	-0.048	4.763	2.238	0.056	2.238
Neely-Weller bench.	4.692	-0.022	4.692	2.312	0.358	2.284
MSIH(4,0)-VAR(1)	4.691	0.017	4.691	2.228	0.002	2.228
PT – BIC selection	4.653	0.033	4.653	2.262	0.015	2.262
AR - BIC selection	4.698	-0.002	4.698	2.300	0.363	2.271
	Panel B: Other benchmarks					
Baseline PT – Adj. R^2 selection	4.639	-0.352	4.485	2.149	0.192	2.140
All forecast functions	5.234	-1.340	5.060	3.277	-0.018	3.277
Equally weighted	4.485	-0.124	4.483	2.093	0.400	2.054
Inv. MSFE-weighted	4.100	-0.146	4.097	1.924	0.358	1.890

Table 8

Summary Statistics for Recursive Mean-Variance Portfolio Weights under a Variety of Forecasting Models/Combination Schemes

The table reports summary statistics for the weights solving the one-month forward mean-variance portfolio problem:

$$\max_{\mathbf{w_t}} E_t[W_{t+1}] - 1/2\,\lambda\, Var_t[W_{t+1}],$$

where W_{t+1} is end-of period wealth and λ is a coefficient of (absolute) risk aversion that trades-off mean and variance. The problem is solved recursively over the period 1985:01 – 2004:11 using in each month updated parameter estimates. The table shows means and standard deviations for recursive portfolio weights. Short-sales are ruled out. Results are shown for a few alternative values of the risk-aversion coefficient, λ.

		Stocks		Bonds		Cash	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
λ = 0.2	MSIH(4,0)-VAR(1)	0.709	0.453	0.166	0.373	0.125	0.328
	MSIAH(3,1)	0.401	0.486	0.259	0.434	0.340	0.469
	VAR(1)	0.407	0.455	0.345	0.454	0.249	0.417
	Neely-Weller benchmark	1.000	0.000	0.000	0.000	0.000	0.000
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.000	0.000	0.000	0.000	0.000	0.000
		Forecast combinations					
	Equally weighted	0.613	0.487	0.230	0.420	0.156	0.362
	Inverse MSFE weights	0.593	0.491	0.240	0.427	0.167	0.373
λ = 0.5	MSIH(4,0)-VAR(1)	0.691	0.453	0.171	0.369	0.139	0.336
	MSIAH(3,1)	0.383	0.475	0.255	0.426	0.362	0.469
	VAR(1)	0.279	0.366	0.399	0.450	0.322	0.429
	Neely-Weller benchmark	1.000	0.000	0.000	0.000	0.000	0.000
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.000	0.000	0.000	0.000	0.000	0.000
		Forecast combinations					
	Equally weighted	0.612	0.488	0.230	0.421	0.158	0.364
	Inverse MSFE weights	0.589	0.491	0.241	0.425	0.169	0.374
λ = 1	MSIH(4,0)-VAR(1)	0.666	0.454	0.181	0.373	0.154	0.345
	MSIAH(3,1)	0.353	0.456	0.260	0.423	0.387	0.470
	VAR(1)	0.170	0.256	0.434	0.445	0.396	0.438
	Neely-Weller benchmark	0.759	0.149	0.241	0.068	0.000	0.000
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.000	0.000	0.000	0.000	0.000	0.000
		Forecast combinations					
	Equally weighted	0.607	0.486	0.230	0.416	0.163	0.365
	Inverse MSFE weights	0.581	0.489	0.246	0.424	0.173	0.376
λ = 2	MSIH(4,0)-VAR(1)	0.595	0.447	0.191	0.367	0.214	0.365
	MSIAH(3,1)	0.288	0.405	0.271	0.414	0.441	0.464
	VAR(1)	0.081	0.134	0.413	0.431	0.505	0.430
	Neely-Weller benchmark	0.720	0.039	0.280	0.039	0.000	0.000
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.000	0.000	0.000	0.000	0.000	0.000
		Forecast combinations					
	Equally weighted	0.593	0.482	0.235	0.411	0.173	0.370
	Inverse MSFE weights	0.563	0.490	0.255	0.425	0.182	0.380

Table 9

Summary Statistics for Recursive Mean-Variance Portfolio Performances under a Variety of Forecasting Models/Combination Schemes

The table reports summary statistics for the 1-month portfolio return based on weights that solve the one-month mean-variance portfolio problem:

$$\max_{\mathbf{w}_t} E_t[W_{t+1}] - 1/2\,\lambda\, Var_t[W_{t+1}],$$

where W_{t+1} is end-of period wealth and λ is the coefficient of (absolute) risk aversion. The problem is solved recursively over the period 1985:01 – 2004:11 using in each month updated parameter estimates. Short-sales are ruled out. Results are shown for a few alternative values of the risk-aversion coefficient, λ. Mean-variance' is the sample realized value of the mean-variance objective. Boldfaced values indicate the best performing model.

		Mean	Median	90% lower bound	90% lower bound	Sharpe ratio	Mean-variance
λ = 0.2	50% stocks, 50% bonds	1.010	1.015	0.954	1.050	0.153	1.009
	1/N benchmark	1.008	1.011	0.970	1.035	0.163	1.007
	MSIH(4,0)-VAR(1)	1.012	1.011	0.961	1.050	0.230	1.012
	MSIAH(3,1)	1.008	1.005	0.961	1.046	0.128	1.008
	VAR(1)	1.011	1.006	0.961	1.048	0.191	1.010
	Neely-Weller benchmark	1.012	1.016	0.955	1.053	0.197	1.011
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.012	1.016	0.955	1.053	0.197	1.011
		Forecast combinations					
	Equally weighted	1.017	1.016	0.974	1.052	0.345	1.017
	Inverse MSFE weights	**1.017**	**1.016**	0.977	1.052	**0.353**	**1.017**
λ = 0.5	50% stocks, 50% bonds	1.010	1.015	0.954	1.050	0.153	1.009
	1/N benchmark	1.008	1.011	0.970	1.035	0.163	1.007
	MSIH(4,0)-VAR(1)	1.012	1.010	0.961	1.048	0.221	1.011
	MSIAH(3,1)	1.008	1.004	0.963	1.046	0.137	1.008
	VAR(1)	1.010	1.005	0.965	1.046	0.183	1.009
	Neely-Weller benchmark	1.012	1.016	0.955	1.053	0.197	1.011
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.012	1.016	0.955	1.053	0.197	1.011
		Forecast combinations					
	Equally weighted	1.017	1.016	0.974	1.052	0.346	1.016
	Inverse MSFE weights	**1.017**	**1.016**	0.977	1.052	**0.354**	**1.016**
λ = 1	50% stocks, 50% bonds	1.010	1.015	0.954	1.050	0.153	1.008
	1/N benchmark	1.008	1.011	0.970	1.035	0.163	1.007
	MSIH(4,0)-VAR(1)	1.011	1.010	0.961	1.046	0.211	1.010
	MSIAH(3,1)	1.008	1.004	0.963	1.046	0.138	1.007
	VAR(1)	1.009	1.005	0.973	1.039	0.157	1.007
	Neely-Weller benchmark	1.012	1.016	0.955	1.053	0.197	1.011
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.012	1.016	0.955	1.053	0.197	1.011
		Forecast combinations					
	Equally weighted	1.017	1.016	0.974	1.052	0.349	1.015
	Inverse MSFE weights	**1.017**	**1.016**	0.977	1.052	**0.355**	**1.016**
λ = 2	50% stocks, 50% bonds	1.010	1.015	0.954	1.050	0.153	1.006
	1/N benchmark	1.008	1.011	0.970	1.035	0.163	1.006
	MSIH(4,0)-VAR(1)	1.010	1.008	0.963	1.044	0.199	1.007
	MSIAH(3,1)	1.008	1.004	0.970	1.044	0.130	1.005
	VAR(1)	1.007	1.005	0.976	1.030	0.118	1.005
	Neely-Weller benchmark	1.011	1.016	0.955	1.052	0.181	1.007
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.012	1.016	0.955	1.053	0.197	1.011
		Forecast combinations					
	Equally weighted	1.017	**1.016**	0.977	1.052	0.356	**1.014**
	Inverse MSFE weights	**1.017**	1.015	0.977	1.052	**0.359**	1.014

Table 10

Sub-Sample Results for Recursive Mean-Variance Portfolio Performances

The table reports summary statistics for the 1-month portfolio return based on weights that solve the one-month mean-variance portfolio problem. Results are broken down for four distinct periods: Jan. 1985 – Dec. 1989, Jan. 1990 – Dec. 1994, Jan. 1995 – Dec. 1999, and Jan. 2000 – Dec. 2004. Short-sales are ruled out. Results are shown for two alternative values of the risk-aversion coefficient, λ. Mean-variance' is the sample realized value of the mean-variance objective.

		λ = 0.5				λ = 1			
		Mean	**Median**	**Sharpe ratio**	**Mean-variance**	**Mean**	**Median**	**Sharpe ratio**	**Mean-variance**
1985 - 1989	50% stocks, 50% bonds	1.012	1.015	0.172	1.010	1.012	1.015	0.172	1.009
	1/N benchmark	1.010	1.012	0.196	1.009	1.010	1.012	0.196	1.008
	MSIH(4,0)-VAR(1)	1.016	1.015	0.334	1.015	1.016	1.015	0.323	1.014
	MSIAH(3,1)	1.011	1.007	0.178	1.010	1.011	1.007	0.181	1.009
	VAR(1)	1.016	1.008	0.345	1.015	1.014	1.008	0.318	1.013
	Neely-Weller benchmark	1.014	1.018	0.227	1.013	1.014	**1.018**	0.227	1.012
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.014	1.018	0.227	1.013	1.014	**1.018**	0.227	1.012
		Forecast combinations							
	Equally weighted	1.018	1.017	0.284	1.019	1.018	1.018	0.229	1.017
	Inverse MSFE weights	**1.021**	**1.018**	**0.409**	**1.020**	**1.021**	1.017	**0.413**	**1.019**
1990 – 1994	50% stocks, 50% bonds	1.008	1.012	0.141	1.012	1.008	1.012	0.141	1.007
	1/N benchmark	1.007	1.009	0.152	1.006	1.007	1.009	0.152	1.006
	MSIH(4,0)-VAR(1)	1.009	1.010	0.206	1.009	1.009	1.010	0.214	1.009
	MSIAH(3,1)	1.009	1.004	0.204	1.009	1.009	1.004	0.210	1.008
	VAR(1)	1.008	1.004	0.169	1.007	1.008	1.004	0.183	1.007
	Neely-Weller benchmark	1.010	**1.013**	0.196	1.009	1.010	1.013	0.196	1.009
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.010	**1.013**	0.196	1.009	1.010	1.013	0.196	1.009
		Forecast combinations							
	Equally weighted	1.010	1.013	0.257	1.011	1.010	**1.013**	0.261	1.010
	Inverse MSFE weights	**1.015**	1.013	**0.431**	**1.015**	**1.015**	1.013	**0.431**	**1.014**
1995 - 1999	50% stocks, 50% bonds	1.019	1.028	0.378	1.018	1.019	1.028	0.378	1.017
	1/N benchmark	1.014	1.020	0.390	1.014	1.014	1.020	0.390	1.013
	MSIH(4,0)-VAR(1)	1.021	1.024	0.532	1.020	1.019	1.018	0.501	1.018
	MSIAH(3,1)	1.010	1.004	0.203	1.010	1.010	1.004	0.200	1.009
	VAR(1)	1.013	1.005	0.291	1.013	1.011	1.005	0.246	1.010
	Neely-Weller benchmark	1.021	**1.030**	0.427	1.020	1.021	1.030	0.427	1.019
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.021	**1.030**	0.427	1.020	1.021	1.030	0.427	1.019
		Forecast combinations							
	Equally weighted	1.021	1.030	0.427	1.020	1.021	**1.030**	0.427	1.019
	Inverse MSFE weights	**1.026**	1.029	**0.624**	**1.025**	**1.026**	1.029	**0.624**	**1.025**
2000 - 2004	50% stocks, 50% bonds	1.001	1.001	-0.042	1.000	1.001	1.001	-0.042	0.999
	1/N benchmark	1.001	1.007	-0.051	1.001	1.001	1.007	-0.051	1.000
	MSIH(4,0)-VAR(1)	1.000	1.004	-0.076	0.999	1.000	1.004	-0.076	0.998
	MSIAH(3,1)	1.003	1.002	-0.002	1.002	1.003	1.002	-0.001	1.001
	VAR(1)	1.002	1.002	-0.018	1.001	1.001	1.003	-0.041	0.999
	Neely-Weller benchmark	1.002	1.011	-0.020	1.001	1.002	1.011	-0.020	1.000
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.002	1.011	-0.020	1.001	1.002	1.011	-0.020	1.000
		Forecast combinations							
	Equally weighted	1.005	**1.014**	0.045	1.003	1.002	1.011	0.054	1.000
	Inverse MSFE weights	**1.007**	1.011	**0.078**	**1.006**	**1.007**	**1.011**	**0.087**	**1.005**

Table 11

Recursive Mean-Variance Portfolio Performances under Transaction Costs

The table reports summary statistics for the 1-month portfolio return based on weights that solve the one-month mean-variance portfolio problem. Short-sales are ruled out. Results are shown for two alternative values of the risk-aversion coefficient, λ. Mean-variance' is the sample realized value of the mean-variance objective. Transaction costs are modeled as the function:

$$tc_t = \phi_p \sum_{j=1}^{n} |\hat{w}_t^j - \hat{w}_{t-1}^j| + \phi_f I_{(\max_j |\hat{w}_t^j - \hat{w}_{t-1}^j|) \neq 0}$$

where ϕ_p is a proportional transaction cost (set to 0.25 percent) and ϕ_f is a fixed cost that has to be paid when transactions occur (ϕ_f is set to 0.01 percent).

		Mean	Median	90% lower bound	90% lower bound	Sharpe ratio	Mean-variance
λ = 0.5	50% stocks, 50% bonds	1.010	1.015	0.954	1.050	0.153	1.009
	1/N benchmark	1.008	1.011	0.970	1.035	0.163	1.007
	MSIH(4,0)-VAR(1)	1.010	1.009	0.958	1.047	0.176	1.009
	MSIAH(3,1)	1.006	1.003	0.963	1.043	0.069	1.005
	VAR(1)	1.007	1.004	0.961	1.042	0.104	1.006
	Neely-Weller benchmark	1.012	1.016	0.955	1.053	0.197	1.010
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.012	1.016	0.955	1.053	0.197	1.011
		Forecast combinations					
	Equally weighted	**1.014**	**1.017**	0.952	1.055	0.236	1.012
	Inverse MSFE weights	1.013	1.013	0.972	1.050	**0.280**	**1.014**
λ = 1	50% stocks, 50% bonds	1.010	1.015	0.954	1.050	0.153	1.008
	1/N benchmark	1.008	1.011	0.970	1.035	0.163	1.007
	MSIH(4,0)-VAR(1)	1.009	1.008	0.958	1.045	0.162	1.008
	MSIAH(3,1)	1.006	1.003	0.961	1.043	0.069	1.004
	VAR(1)	1.006	1.004	0.971	1.036	0.086	1.005
	Neely-Weller benchmark	1.012	1.016	0.955	1.053	0.197	1.010
	Pesaran-Timmermann ($\bar{R}^2$ selection)	1.012	1.016	0.955	1.053	0.197	1.010
		Forecast combinations					
	Equally weighted	1.013	**1.016**	0.954	1.053	0.241	1.011
	Inverse MSFE weights	**1.014**	1.013	0.973	1.051	**0.281**	**1.013**